

TECHNICAL DIAGNOSTIC SYSTEM IN THE MAINTENANCE OF TURBOMACHINERY FOR AMMONIA SYNTHESIS IN THE PROCESS INDUSTRIES

Natalia Nikolova^{a,b*}, Kaoru Hirota^c, Krasimir Kolev^d, Kiril Tenekedjiev^{a,b}

^a Australian Maritime College, University of Tasmania, 1 Maritime Way, 7250 Launceston TAS (Australia); Natalia.Nikolova@utas.edu.au; Kiril.Tenekedjiev@utas.edu.au

^b Nikola Vaptsarov Naval Academy, 73 V. Drumev Str., 9026 Varna, Bulgaria, natalianik@gmail.com; Kiril.Tenekedjiev@fulbrightmail.org

^c School of Automation, Beijing Institute of Technology, No.5 Zhongguancun South Street, Haidian District, Beijing 100081, China; hirota@bit.edu.cn, molimhirota@mtd.biglobe.ne.jp

^d Department of Medical Biochemistry, Semmelweis University, 1085 Budapest, Üllői út 26. Hungary, Kolev.Krasimir@med.semmelweis-univ.hu

* Corresponding author

Abstract

Technical maintenance of machines and equipment in processing industry requires elaborate technical diagnostics systems to recognize the current state and forecast their future state. Creating such a system is a complex task due to multiple factors, with aging in aggressive exploitation environment being an important one. Statistical pattern recognition systems are very suitable to solve problems of technical diagnostics as they produce quantitative estimates of the states. We present the use of a hybrid Bayesian pattern recognition classifier that utilizes statistical and fuzzy paradigms and expresses the measurement information with four types of features (discrete, pseudo-discrete, multi-normal and independent continuous). It uses frequentist and subjective information (from training samples and expert opinion respectively) to identify the unknown parameters of the conditional likelihood density functions of each technical state. We discuss possible sources to collect learning information, and different methods to represent it. The classifier uses three different methods for parameter estimation of the conditional likelihood densities using data fusion. The classification is realised as a discriminant non-linear machine, which incorporates fuzzy approaches at different levels. We develop a novel algorithm for fault prediction without dynamic learning with four possible types of answers. A detailed example of technical diagnostics system for classification and prediction of states of turbomachinery for ammonia synthesis is presented. For the journal bearing diagnostics, we introduce modification of the hybrid Bayesian classifier using pseudo-priors to incorporate rule-based knowledge and improve the classification.

Keywords: hybrid Bayesian classifier, pseudo-discrete features, fuzzy parameter estimation, backward discriminant functions, aging, pseudo-priors, ammonia synthesis

1. Introduction

Proper and consistent technical maintenance procedures are the way to increase profitability, quality and reliability in exploitation of complex machines and equipment. These procedures are environmentally-friendly and allow to minimize the time that the equipment stays out of operation (due to break downs and long-lasting repairs). A good example is turbo machinery in chemical processing industry. Technical maintenance is a resource-demanding activity (sometimes reaching 1.5-2 times the investment in complex production equipment), which is another reason why the choice of technical maintenance strategy is of crucial importance. There are four main technical maintenance strategies – event-driven, time-driven, actual-state-driven and mixed strategy. Each strategy has its implementation areas depending on the requirements of the machinery.

The event-driven maintenance assumes that technical service to the equipment is provided when a technical failure occurs. That strategy is suitable for home appliances or for high-reliability

equipment such as sophisticated computers or other hi-tech machinery, where moral aging occurs before mechanical aging.

The time-driven maintenance assumes that technical servicing and repairs of the equipment are made at regular, pre-defined time intervals (Zavoichinskaya, Zavoichinskii, 2006). It is the widest spread approach in technical maintenance. It is useful when the equipment can be restored to its original condition at every preventive maintenance in case the failure rate is increasing. Otherwise this strategy becomes counterproductive (Ebeling, 2010, pp. 237-247). In this strategy, after completion of a given time of operation the equipment goes for scheduled maintenance or repair and stays inoperable for a given duration of time. This is an expensive approach since regardless of how precisely the time interval between repairs ΔT was identified, it will practically always differ from the actual interval ΔT^o between the last servicing of the equipment and the occurrence of a technical failure. If $\Delta T^o > \Delta T$ then functional segments of the equipment are being replaced and technical maintenance was performed too early. If $\Delta T > \Delta T^o$ then the technical failure may also lead to expensive non-scheduled repairs. The strength of this strategy is the ability to schedule the maintenance periods in the operational cycle of the equipment, on the expense of some potential financial and exploitation risks.

The actual-state-driven maintenance (also called condition-based maintenance, see (Hadden et al, 2000) may lead to major economic benefits since maintenance is only performed when technical fault emerges (or is detected). This strategy requires constant measurement of exploitation parameters and constant information about the technical status of equipment. It is also necessary to follow the patterns of how performance parameters change to make proper forecasts for eventual failures (hence schedule timely repairs).

The fourth strategy is a mixed one, which relies on estimation of the technical state of the equipment in regular intervals of time ΔT^g as well as on needs for repairs dictated by the deteriorating exploitation characteristics. The repair is conducted only after deterioration of performance is proven. The effectiveness of this approach is present if $\Delta T^g \ll E[\Delta T^o]$ and the variance $V[\Delta T^o]$ is low. Those two conditions guarantee that there will be a maintenance activity before the failure occurs. In this case, it is crucial to be able to forecast the future state of the equipment within the interval ΔT^g , which implies that the probability to observe failure till the next time we estimate technical state would be low.

The types of failures that are most typical for a given set of equipment also impacts the choice of maintenance strategy. Equipment that breaks unexpectedly without visible or detectable indication requires to use the second strategy. On the other hand, the third and fourth strategies are economically sustainable for gradually developing faults, i.e. when the exploitation of the system in deteriorated technical state and the deterioration of production quality leads to changes that are initially negligibly small but later become explicitly evident and may lead to severe failures and long-lasting repairs. The aging of equipment in process industry is an example of such situation.

The effectiveness of the third and fourth strategies depends on the presence of a technical diagnostics system (TDS). Technical diagnostics has the task to recognize the current state and forecast the future state of technical systems (Iserman, 2006; Vachtsevanos et al., 2006). The most effective are those methods that give a quantitative estimate of the state. Those may be incorporated into a general scheme for technical maintenance management. They have a positive effect on the repairs and sometimes technical diagnostics is perceived as actual-state-feedback in the control system of complex equipment. Designing a TDS in processing industry is not trivial because of the following peculiarities:

- 1) The connection between the states and the values of the diagnostic features is non-deterministic and unreliable. As a rule, many low-informative features contain the information about most of the faults.
- 2) It is impossible to describe the objects of TDS with differential equations due to complex deterministic or stochastic feedbacks. Differential equations can be used to described only

unrealistically over-simplified models of equipment (Marinov, et al., 2014; Marinov et al., 2016)

- 3) The presence of a great number of features, usually multi-dimensional and bound by complex correlations
- 4) The qualitative changes in the state of the TDS objects occur continuously due to aging and not in a deterministic fashion
- 5) The objects of TDS (e.g. powerful multi-section turbines and compressors, rotational and piston machinery, large heat exchangers, etc.) are very expensive and their projected reliability is very high, therefore it is very hard to find training samples
- 6) Conducting active experiments during the presence of faults is practically impossible
- 7) Due to high risk of failure of expensive equipment segments and disruption of working parameters during continuous operations
- 8) The most important thing about TDS is to detect the tendency or the trend, i.e. to give quantitative estimation of the technical state
- 9) The responsibility for fault-free operation is carried by engineers and technicians, so that it is impossible to completely trust the system, instead it is necessary to visualize the classification and forecasting. The introduction of a TDS is a major organizational change in technical maintenance, which is always a challenge (Nilsen, 2006)
- 10) Most equipment in the processing industry operates in an aggressive production environment, which further accelerates the aging process

Technical diagnostic systems are also known as fault detection and isolation systems (FDI) (as tackled by the control engineering community (Patton et al., 2000, pp. 15-49)). There are several types of such systems. Model based FDIs utilize a given model of the system, which allows to make predictions for the analysed quantities and parameters during the current operating regime and decide about occurrence of faults (Heng et al., 2009). The accuracy of the utilized model is a key issue since it predefines the accuracy of the analysis of the system behaviour (Rajamani, Cheded, 2017). The discrepancy between the true measurements and the ones forecasted by the model define the so-called residuals. These are compiled in a vector and serve to detect various types of faults. If a given snap shot of observations at a given time indicates the presence of a given fault, the residuals are treated as signatures of the fault (Briones et al., 2008) (Bregon et al., 2008).

There are many methods for model-based diagnostics, including parity-space approach, observer-based approach (Ding, 2008, pp. 21-45), qualitative fault signatures (Mosterman, Biswas, 1999) and possible conflicts (PCs) (Pulido, Alonso-Gonzalez, 2004), analytic redundancy relation (ARR) (Samantaray et al., 2006) (Bregon et al., 2008). The greatest advantage of such model based FDIs is that they attempt to find all faults in the system at the same time. These methods often assume there is no noise in the system, which has been postulated by works both in the control community (Patton et al., 2000, 2-13) and in the artificial intelligence community (Cordier et al., 2004). Alternative view on the system containing stochastic parts and the observations having probability distributions rather than exact values is suggested by (Krysander, Nyberg, 2008). However, the construction of such models requires a great deal of learning information about the physical process under analysis. Therefore, such systems are successful only when there is a small number of highly informative features, indicating various faults.

Another type of FDIs are the signal processing FDIs, where series of experiments are conducted to diagnose faults based on the results of those experiments using mathematical or statistical operations (Bahrapour et al., 2011). These systems also require large sets of learning information, which is often missing. They are rarely known to generate probabilities for the faults, instead those methods only identify the possible fault existent in the system.

A third type of systems are the artificial intelligence FDIs. The essence of those methods is some form of a classifier, based on various components such as artificial neural networks, expert systems, non-parametric and semi-parametric systems (K-nearest neighbour, support vector machines (Palade et al., 2002), Parsen windows, etc.), or a Bayesian classifier (Rajamani,

Cheded, 2017). A typical characteristic of such systems is the usage of many low-informative features, where each possible fault is diagnosed by various models (i.e. there is a separate part of the system that deals with each fault).

The strengths of Bayesian classifiers in machine fault diagnostics are the following:

- They allow to conveniently combine all types of learning information from expert knowledge or from training samples
- They find posterior probabilities for the technical status of an object in any moment of time. Traditionally, they analyse faults as separate classes of state, which also includes the no-fault class. Often, if there is a situation where more than one fault occurs at the same time, then these situations are excluded as being very rare in this context. The result is usually a single vector of posterior probabilities, describing all states simultaneously.
- The presence on a single vector of posterior probabilities leads to the following:
 - o Clear picture of the status of the system in each moment of time
 - o Possibility to create a decision-making system to control the maintenance, where the uncertainty in the system is evaluated by the Bayesian classification system (Cullum, Tenekedjiev, 2017)
 - o It is possible and convenient to make forecasts using statistical information
- The probability for error during classification is theoretically the lowest for Bayesian classifiers. As it is proven in Duda et al. (2001, pp. 45-46), the use of the Bayesian rule (i.e. that the observation is classified into the class with maximum posterior probability) maximizes the chance of being correct for any conditional likelihoods.
- There are well-developed methods for feature selection and feature extraction, which are part of statistical pattern classification
- They allow to run the system and make classifications fast while allowing the system to use left-learning procedures to upgrade its diagnostic capabilities
- There are well-developed methods for performance estimation that include many parameters, such as confusion matrix (indicating the percentage of the i -th class classified into the j -th class), certainty matrix (average percentage to classify the i -th class in class j) and doubt matrix (average percentage of the correct class if classification is wrong and average percentage of the first incorrect class if classification was correct) (Nikolova et al., 2013)
- The learning sample for classification, learning and recognition can transfer from one data base to another very flexibly (Tenekedjiev et al., 2006)

The most elaborate TDS are the pattern classification systems. There are different types of pattern recognition (PR) and pattern classification approaches: syntactic PR, structural PR, neural networks PR (Hambebo et al., 2013), template matching PR (Jain et al., 2000), fuzzy PR (Klir, Yuan, 1995, pp. 357-374), and statistical PR. The first four of them are qualitative-oriented rather than quantitative. The fuzzy PR tries to emulate human behavior. Some basic human classification techniques have been identified by Zadeh (1977), to develop a formal fuzzy PR system. The different types of uncertainty (that can influence a PR problem) are investigated by Klir (2006). The possibility theory introduced in (Zadeh, 1978) allows reasoning with uncertain, imprecise, ambiguous, and/or conflicting vague knowledge. The possibilities and advantages of combining fuzzy set theory and PR are outlined by Pedrycz (1997) and Tonon, (1999). In the statistical PR, observations are represented by a set of measured values (called features) organized in multi-dimensional feature vectors. The objective is to select those features, and/or extract new ones, which after appropriate transformation allow the observation vectors belonging to different categorical classes to occupy compact and mainly disjoint regions in the multidimensional space. The separation of those regions is made according to conditional distributions of each of the categorical class feature vectors (Jain et al., 2000). Widely used is the Bayesian technique (based on

the Bayes theorem, see Pratt et al. (2008)), where the recognition is made according to the object's maximum probability to belong to each of the classes, given a certain observation vector (Duda et al., 2001). Apart from the latter, all other techniques are qualitative-oriented rather than quantitative.

In Appendix A, we outline the hybrid Bayesian PR classifier of Tenekedjiev et al. (2006). There are several reasons why that classifier is suitable for TDS in processing industry:

- a) New data formats are employed for data presentation allowing the observation vector to operate on four types of features: discrete, pseudo-discrete, multi-normal and independent continuous.
- b) The Bayesian approach has a minimal overall error of classification from all other approaches as discussed by Duda et al. (2001)
- c) It is possible to train the selected classifier utilizing the frequentist and the subjective definition of probabilities and even to combine them.
- d) The proposed classifier uses fuzzy techniques in the data presentation, in the learning data fusion, in the learning phase and in the recognition phase.
- e) The posterior probabilities for each state are calculated and they can in turn be the input to a general decision support system about the maintenance of technical objects (see Gammack et al. 2011)
- f) This classifier adapts very well on one hand to the changes in the object (caused by aging) and on the other hand – to the changing quality and quantity of the learning information.

Other examples for hybrid approaches can be found in Hadjiski and Doukovska (2012) (hybrid neuro-fuzzy network classifiers), and Tonon et al. (1999) (for hybrid fuzzy and statistical classifier).

The objective of this paper is to discuss the sources and methods of learning information for pattern recognition and then present elaborate methods for parameter estimation of conditional likelihood density functions using four types of feature descriptions. As a result, we will present procedures to build non-linear machine of discriminant functions for multi-class classification of technical state of equipment. We will apply the hybrid Bayesian classification approach for technical diagnostics of turbomachinery used for ammonia synthesis. In Section 2 we discuss the various sources and methods for collection of learning information for PR. In Section 3 we analyse different methods to process learning information to estimate parameters of the conditional likelihood density functions in the hybrid Bayesian PR classifier using three methods. In Section 4, we formulate the non-linear machine of discriminant functions for classification and transform the discriminant functions into posterior probabilities, assessed by that hybrid Bayesian PR classifier. Section 5 gives a case study for the application of the hybrid Bayesian PR classifier in ammonia production.

2. Sources of learning information used in the hybrid Bayesian classifier

For the construction of a Bayesian classifier of type (A.1) it is necessary to estimate the parameters

$$p_{j,i}^k, \text{ for } i=1,2,\dots, a; j=1,2,\dots, b_i; k=1,2,\dots, c, \quad (1)$$

$$q_{j,i}^k, \text{ for } i=1,2,\dots, t; j=1,2,\dots, h_i; k=1,2,\dots, c, \quad (2)$$

$$\mu_i^k, \text{ for } i=1,2,\dots, e; k=1,2,\dots, c, \quad (3)$$

$$\sigma_i^k, \text{ for } i=1,2,\dots, e; k=1,2,\dots, c, \quad (4)$$

$$r_{i,j}^k, \text{ for } i=1,2,\dots, e-1; j=i+1, i+2, \dots, e; k=1,2,\dots, c \quad (5)$$

$$z_{\alpha_j,i}^k, \text{ for } i=1,2,\dots, g; j=1,2,\dots, m; k=1,2,\dots, c. \quad (6)$$

For the conditional likelihood densities' estimation, a hybrid parametric-nonparametric approach is employed like the recently developed semi-parametric density estimation approach

(Hoti and Holmstrom, 2003). The entities for the conditional likelihood densities' estimation are called parameters, just for the sake of simplicity.

Two main sources of learning information are used for parameter estimation: training samples and experts estimates for the conditional likelihood densities' parameters.

The training sample $\chi_k = \{\bar{x}_l^k - L_l^{k(x)} \mid l=1,2,\dots,n_k\}$ for class k contains n_k couples. Each couple consists of an observation vector \bar{x}_l^k which belongs to the set of class k examples with membership $L_l^{k(x)}$. Here,

$$\bar{x}_l^k = \left(\bar{x}_l^{k(d)T}, \bar{x}_l^{k(p)T}, \bar{x}_l^{k(c)T}, \bar{x}_l^{k(i)T} \right)^T \text{ for } j=1,2,\dots,n_k; k=1,2,\dots,c \quad (7)$$

$$\bar{x}_l^{k(d)} = \left(x_{l,1}^{k(d)}, x_{l,2}^{k(d)}, \dots, x_{l,a}^{k(d)} \right)^T \text{ for } l=1,2,\dots,n_k; k=1,2,\dots,c \quad (8)$$

$$\bar{x}_l^{k(p)} = \left(x_{l,1}^{k(p)}, x_{l,2}^{k(p)}, \dots, x_{l,t}^{k(p)} \right)^T \text{ for } l=1,2,\dots,n_k; k=1,2,\dots,c \quad (9)$$

$$\bar{x}_l^{k(c)} = \left(x_{l,1}^{k(c)}, x_{l,2}^{k(c)}, \dots, x_{l,e}^{k(c)} \right)^T \text{ for } l=1,2,\dots,n_k; k=1,2,\dots,c \quad (10)$$

$$\bar{x}_l^{k(i)} = \left(x_{l,1}^{k(i)}, x_{l,2}^{k(i)}, \dots, x_{l,g}^{k(i)} \right)^T \text{ for } l=1,2,\dots,n_k; k=1,2,\dots,c \quad (11)$$

The fuzzy confidence coefficient of the expert $L_j^{k(x)} \in (0; 1]$ facilitates the data fusion of the learning information and allows learning not only from measurement of S itself, but also from observations of similar objects of S .

The expert information for class k can be divided into four groups: the conditional probabilities of the discrete features, the conditional probabilities of the pseudo-discrete features, the parameters (mean values, standard deviations, and correlation coefficients) of the multi-normal features, and the conditional quantiles of the independent-continuous features of S . Each group exists in its entirety or is absent altogether. All parameters can be directly or indirectly estimated by experts as suggested by Brown (2005).

The expert information about the conditional probabilities of the discrete features

$E_k^{(d)} = \{\bar{E}_i^{k(d)} - L_i^{k(d)} \mid i=1,2,\dots,a\}$ for class k contains a couples. Each couple consists of expert estimate $\bar{E}_i^{k(d)}$ for the conditional probabilities of df_i and of a coefficient $L_i^{k(d)}$ which measures the confidence of the expert about the estimates in $\bar{E}_i^{k(d)}$. Here

$$\bar{E}_i^{k(d)} = \left(p_{1,i}^{k(e)}, p_{2,i}^{k(e)}, \dots, p_{b,i}^{k(e)} \right) \text{ for } i=1,2,\dots,a; k=1,2,\dots,c \quad (12)$$

The fuzzy confidence coefficient for the df_i conditional probabilities' estimates $L_i^{k(d)} \in (0; 1]$ increases with the confidence of the expert.

The expert information about the conditional probabilities of the pseudo-discrete features

$E_k^{(p)} = \{\bar{E}_i^{k(p)} - L_i^{k(p)} \mid i=1,2,\dots,t\}$ for class k contains t couples (for further discussion on pseudo discrete features see (Fatichah et al., 2015)). Each couple consists of expert estimate $\bar{E}_i^{k(p)}$ for the conditional probabilities of pf_i and of a coefficient $L_i^{k(p)}$ which measures the confidence of the expert about the estimates in $\bar{E}_i^{k(p)}$. Here

$$\bar{E}_i^{k(p)} = \left(q_{1,i}^{k(e)}, q_{2,i}^{k(e)}, \dots, q_{h,i}^{k(e)} \right) \text{ for } i=1,2,\dots,t; k=1,2,\dots,c \quad (13)$$

The fuzzy confidence coefficient for the pf_i conditional probabilities' estimates $L_i^{k(p)} \in (0; 1]$ increases with the confidence of the expert.

The expert information about the distribution of the multi-normal features

$E_k^{(c)} = M_k \cup D_k \cup R_k$ for class k contains three sub-groups:

$$M_k = \left\{ \mu_i^{k(e)} - L_i^{k(\mu)} \mid i = 1, 2, \dots, e \right\} \quad (14)$$

$$D_k = \left\{ \sigma_i^{k(e)} - L_i^{k(\sigma)} \mid i = 1, 2, \dots, e \right\} \quad (15)$$

$$R_k = \left\{ r_{i,j}^{k(e)} - L_{i,j}^{k(r)} \mid i = 1, 2, \dots, e-1; j = i+1, i+2, \dots, e \right\} \quad (16)$$

Here $\mu_i^{k(e)}$ and $\sigma_i^{k(e)}$ are the conditional expert's estimates for the mean value and the standard deviation of cf_i , $r_{i,j}^{k(e)}$ is the conditional expert's estimate for the correlation coefficient between cf_i and cf_j , whereas the coefficients $L_i^{k(\mu)}$, $L_i^{k(\sigma)}$, and $L_{i,j}^{k(r)}$ measure the confidence of the expert about the estimates $\mu_i^{k(e)}$, $\sigma_i^{k(e)}$, and $r_{i,j}^{k(e)}$, respectively. The fuzzy confidence coefficients $L_i^{k(\mu)} \in (0; 1]$, $L_i^{k(\sigma)} \in (0; 1]$, and $L_{i,j}^{k(r)} \in (0; 1]$ increase with the confidence of the expert.

The expert information about the conditional quantiles of the independent-continuous features $E_k^{(i)} = \left\{ \bar{E}_i^{k(i)} - L_i^{k(i)} \mid i = 1, 2, \dots, g \right\}$ for class k contains g couples. Each couple consists of expert estimate $\bar{E}_i^{k(i)}$ for the conditional quantiles of if_i and of a coefficient $L_i^{k(i)}$ which measures the confidence of the expert about the estimates in $\bar{E}_i^{k(i)}$. Here

$$\bar{E}_i^{k(i)} = \left(z_{\alpha_1,i}^{k(e)}, z_{\alpha_2,i}^{k(e)}, \dots, z_{\alpha_m,i}^{k(e)} \right) \text{ for } i=1, 2, \dots, g; k=1, 2, \dots, c \quad (17)$$

The fuzzy confidence coefficient for the if_i conditional quantiles' estimates $L_i^{k(i)} \in (0; 1]$ increases with the confidence of the expert.

The count of the fuzzy confidence coefficients is $[c(a+t+g+1.5e+0.5e^2)]$. They are useful for the data fusion of the learning information.

Alternatively, the experts can use preferential equations for elicitation of their parameter estimates. The elicitation process usually takes the form of a series of simple questions in each of which the expert needs to compare two options by her preference. Those options' structure depends on the type of elicited parameter. Quantitative decision analysis (QDA) measures uncertainty of an event through subjective probability elicited by such a series of questions. Each of the discrete probabilities (12) and pseudo-discrete probabilities (13) can be elicited by a procedure advised in (French, 1993, 237-244; French and Insua, 2010, pp. 74-76) because they relate to well defined random events. A similar QDA-procedure exists for elicitation of a subjective quantile from the Cumulative Distribution Function (CDF) of a one-dimensional random variable (r.v.) (see Pratt et al. (2008, pp. 150-155)). If applied, that would result directly in expert estimates for conditional quantiles (17) as well as for the mean values (14). The latter is true because the mean value coincides with the median (the 0,5-quantile) when the marginal distribution of cf_i is normal. The experts' estimates for standard deviations (15) and for correlation coefficients (16) cannot be elicited directly. The reason is that both are moments of the distribution and any statement about their value is unverifiable (French, 1993, 235-236). However, one can obtain QDA-estimates for (15) and (16) by applying the weighted least square procedures proposed by Nikolova et al. (2010). The first procedure elicits the upper and lower quartile for the marginal distribution of cf_i using the above-mentioned quantile elicitation method. It finds the expert estimates (15) as the standard deviation of the normal distribution with the best goodness-of-fit measure. The second procedure elicits the medians of four conditional distributions: the cf_i distribution provided that the cf_j is set on

its lower quartile, the cf_i distribution provided that the cf_j is set on its upper quartile, the cf_j distribution provided that the cf_i is set on its lower quartile, and the cf_j distribution provided that the cf_i is set on its upper quartile. It finds the expert estimates (16) as the correlation coefficient of the bi-normal distribution with the best goodness-of-fit measure.

Any elicited estimates for the parameters of the conditional likelihood functions obtained by QDA-preference equation-procedures are in an interval form usually due to limitations in the discrimination power of the real experts. The tighter the uncertainty interval of an estimate, the larger the fuzzy confidence coefficient of that parameter should be.

3. Parameter estimation in the hybrid Bayesian classifier

The existence of learning information in the form of training samples and in the form of expert estimates enables the system to apply the hybrid frequentist-subjective approach to probabilities.

If n_k is big 'enough' then the conditional likelihood density parameters for class k can be estimated by the maximum-likelihood method:

$$p_{j,i}^k = p_{j,i}^{k(x)} = \sum_{\substack{l=1 \\ x_{l,i}^{k(d)} = df_{j,i}}}^{n_k} L_l^{k(x)} \bigg/ \sum_{l=1}^{n_k} L_l^{k(x)}, \text{ for } i=1,2,\dots, a; j=1,2,\dots, b_i; k=1,2,\dots, c \quad (18)$$

$$q_{j,i}^k = q_{j,i}^{k(x)} = \sum_{l=1}^{n_k} \left[L_l^{k(x)} \mu_{j,i}^{k(p)} \left(x_{l,i}^{k(p)} \right) \right] \bigg/ \sum_{l=1}^{n_k} L_l^{k(x)}, \text{ for } i=1,2,\dots, t; j=1,2,\dots, h_i; k=1,2,\dots, c \quad (19)$$

$$\mu_i^k = \mu_i^{k(x)} = \sum_{l=1}^{n_k} \left[L_l^{k(x)} x_{l,i}^{k(c)} \right] \bigg/ \sum_{l=1}^{n_k} L_l^{k(x)}, \text{ for } i=1,2,\dots, e; k=1,2,\dots, c \quad (20)$$

$$\sigma_i^k = \sigma_i^{k(x)} = \sqrt{(n_k - 1) \sum_{l=1}^{n_k} \left[L_l^{k(x)} \left(x_{l,i}^{k(c)} - \mu_i^{k(x)} \right)^2 \right] \bigg/ \left(n_k \sum_{l=1}^{n_k} L_l^{k(x)} \right)}, \text{ for } i=1,2,\dots, e; k=1,2,\dots, c \quad (21)$$

$$r_{i,j}^k = r_{i,j}^{k(x)} = \frac{(n_k - 1) \sum_{l=1}^{n_k} \left[L_l^{k(x)} \left(x_{l,i}^{k(c)} - \mu_i^{k(x)} \right) \left(x_{l,j}^{k(c)} - \mu_j^{k(x)} \right) \right]}{n_k \sum_{l=1}^{n_k} \left(L_l^{k(x)} \right) \sigma_i^{k(x)} \sigma_j^{k(x)}}, \quad (22)$$

for $i=1,2,\dots, e-1; j=i+1, i+2, \dots, e; k=1,2,\dots, c$

The maximum likelihood estimate of the conditional quantiles for the independent continuous features is not as straightforward as the above formulae. One can get the estimates using the following procedure ($m \times g \times c$)-times for calculating $z_{\alpha_j,i}^k$.

- a. Extract the observations in χ_k for if_i :

$$\chi_i^{k(i)} = \left\{ x_{1,i}^{k(i)}, x_{2,i}^{k(i)}, \dots, x_{n_k,i}^{k(i)} \right\}$$

- b. Sort the elements of $\chi_i^{k(i)}$ by finding $s(1), s(2), \dots, s(n_k)$, which form permutations of $(1,2,3, \dots, n_k)$ such that $x_{s(1),i}^{k(i)} \leq x_{s(2),i}^{k(i)} \leq \dots \leq x_{s(n_k),i}^{k(i)}$

- c. Set $s(0) = 0, x_{s(0),i}^{k(i)} = z_{0,i}, L_{s(0)}^{k(i)} = 0, s(n_k + 1) = n_k + 1, x_{s(n_k+1),i}^{k(i)} = z_{1,i}$, and $L_{s(n_k+1)}^{k(i)} = 0$, where $z_{0,i}$ and $z_{1,i}$ are respectively the minimum and maximum allowed value of if_i . Then form sorted observations in χ_k for if_i : $\chi_i^{k(i,sort)} = \left\{ x_{s(0),i}^{k(i)}, x_{s(1),i}^{k(i)}, x_{s(2),i}^{k(i)}, \dots, x_{s(n_k),i}^{k(i)}, x_{s(n_k+1),i}^{k(i)} \right\}$.

- d. Calculate the vector of the middle points between the elements of $\chi_i^{k(i,sort)}$:

$$\overrightarrow{zT} = \left(xT_{1,i}^{k(i)}, xT_{2,i}^{k(i)}, \dots, xT_{n_k+1,i}^{k(i)} \right) \text{ where}$$

$$xT_{l,i}^{k(i)} = \left(x_{s(l-1),i}^{k(i)} + x_{s(l),i}^{k(i)} \right) / 2, \text{ for } l=1, 2, \dots, n_k+1$$

- e. Calculate the vector of the α values corresponding to the α -quantiles in \overline{zT} :

$$\overline{\alpha T} = (\alpha T_{1,i}, \alpha T_{2,i}, \dots, \alpha T_{n_k+1,i}) \text{ where}$$

$$\alpha T_{l,i} = \sum_{l'=1}^{l-1} L_{s(l')}^{k(x)} / \sum_{l'=1}^{n_k} L_{s(l')}^{k(x)}, \text{ for } l=1, 2, \dots, n_k+1$$

- f. Find $z_{\alpha_j,i}^k = z_{\alpha_j,i}^{k(x)}$ as table lookup from $(\overline{\alpha T}, \overline{zT})$ at the query point α_j :

$$z_{\alpha_j,i}^k = z_{\alpha_j,i}^{k(x)} = \text{interp1}(\overline{\alpha T}, \overline{zT}, \alpha_j), \text{ for } i=1,2,\dots, g; j=1,2,\dots, m; k=1,2,\dots, c \quad (23)$$

The function *interp1* "looks-up" the α_j in $\overline{\alpha T}$, and based upon its location, returns the value $z_{\alpha_j,i}^{k(x)}$

linearly interpolated within the elements of \overline{zT} (Mathworks, 2018).

If n_k is 0 then the conditional likelihood density parameters for class k must be taken from experts estimates:

$$p_{j,i}^k = p_{j,i}^{k(e)}, \text{ for } i=1,2,\dots, a; j=1,2,\dots, b; k=1,2,\dots, c, \quad (24)$$

$$q_{j,i}^k = q_{j,i}^{k(e)}, \text{ for } i=1,2,\dots, t; j=1,2,\dots, h; k=1,2,\dots, c, \quad (25)$$

$$\mu_i^k = \mu_i^{k(e)}, \text{ for } i=1,2,\dots, e; k=1,2,\dots, c, \quad (26)$$

$$\sigma_i^k = \sigma_i^{k(e)}, \text{ for } i=1,2,\dots, e; k=1,2,\dots, c, \quad (27)$$

$$r_{i,j}^k = r_{i,j}^{k(e)}, \text{ for } i=1,2,\dots, e-1; j=i+1, i+2, \dots, e; k=1,2,\dots, c \quad (28)$$

$$z_{\alpha_j,i}^k = z_{\alpha_j,i}^{k(e)}, \text{ for } i=1,2,\dots, g; j=1,2,\dots, m; k=1,2,\dots, c. \quad (29)$$

In all other cases, we have some data in the learning sample and some expert estimates for class k . Then, it is possible to apply pseudo-Bayesian fusion of both the frequentist and the subjective parameter estimates. In that fusion we use the notion that fully reliable expert estimate should be equal to an estimate of approximately 20 observations:

$$p_{j,i}^k = \frac{\sum_{l=1}^{n_k} L_l^{k(x)} p_{j,i}^{k(x)} + 20 L_i^{k(d)} p_{j,i}^{k(e)}}{\sum_{l=1}^{n_k} L_l^{k(x)} + 20 L_i^{k(d)}}, \text{ for } i=1,2,\dots, a; j=1,2,\dots, b; k=1,2,\dots, c, \quad (30)$$

$$q_{j,i}^k = \frac{\sum_{l=1}^{n_k} L_l^{k(x)} q_{j,i}^{k(x)} + 20 L_i^{k(p)} q_{j,i}^{k(e)}}{\sum_{l=1}^{n_k} L_l^{k(x)} + 20 L_i^{k(p)}}, \text{ for } i=1,2,\dots, t; j=1,2,\dots, h; k=1,2,\dots, c, \quad (31)$$

$$\mu_i^k = \frac{\sum_{l=1}^{n_k} L_l^{k(x)} \mu_i^{k(x)} + 20 L_i^{k(\mu)} \mu_i^{k(e)}}{\sum_{l=1}^{n_k} L_l^{k(x)} + 20 L_i^{k(\mu)}}, \text{ for } i=1,2,\dots, e; k=1,2,\dots, c, \quad (32)$$

$$\sigma_i^k = \frac{\sum_{l=1}^{n_k} L_l^{k(x)} \sigma_i^{k(x)} + 20 L_i^{k(\sigma)} \sigma_i^{k(e)}}{\sum_{l=1}^{n_k} L_l^{k(x)} + 20 L_i^{k(\sigma)}}, \text{ for } i=1,2,\dots, e; k=1,2,\dots, c, \quad (33)$$

$$r_{i,j}^k = \frac{\sum_{l=1}^{n_k} L_l^{k(x)} r_{i,j}^{k(x)} + 20 L_i^{k(r)} r_{i,j}^{k(e)}}{\sum_{l=1}^{n_k} L_l^{k(x)} + 20 L_i^{k(r)}}, \text{ for } i=1,2,\dots, e-1; j=i+1, i+2, \dots, e; k=1,2,\dots, c \quad (34)$$

$$z_{\alpha_j,i}^k = \frac{\sum_{l=1}^{n_k} L_l^{k(x)} z_{\alpha_j,i}^{k(x)} + 20 L_i^{k(i)} z_{\alpha_j,i}^{k(e)}}{\sum_{l=1}^{n_k} L_l^{k(x)} + 20 L_i^{k(i)}}, \text{ for } i=1,2,\dots, g; j=1,2,\dots, m; k=1,2,\dots, c. \quad (35)$$

4. Discriminant Function Formation

The direct calculation of the value of the conditional likelihood density if S is in ω_k (A.3) using (A.6), (A.11), (A.21), (A.23) and (A.27) is never a good idea. The reason is that the resulting formula will have multiple positive values less than 1 multiplied. It can often lead to underflow error, or to catastrophic loss of precision of the calculations. Almost as a standard, instead of the conditional likelihood densities (A.3), c number of discriminant functions F_k are calculated. The discriminant function F_k is defined as the dependent on the class k part from the natural logarithm of the posterior probability $P(\omega_k | \vec{x})$:

$$F_k(\vec{x}) = \sum_{\substack{i=1 \\ x_i^{(d)}=df_{j,i}}}^a \ln(p_{j,i}^k) + \sum_{i=1}^t \ln\left(\sum_{j=1}^{h_i} \mu_{j,i}(x_i^{(p)}) q_{j,i}^k\right) + \sum_{i=1}^g p(x_i^{(i)} | \omega_k) + \\ + \left(A\vec{x}^{(c)} - A\vec{\mu}\right)^T W_k \left(A\vec{x}^{(c)} - A\vec{\mu}\right) + \vec{w}_k \left(A\vec{x}^{(c)} - A\vec{\mu}\right) + w_{0,k} \quad , \text{ for } k=1, 2, \dots, c \quad (36)$$

In (36), A is calculated using (A.17), $\vec{\mu}$ is calculated using (A.14), $p(x_i^{(i)} | \omega_k)$ are calculated using (A.27), the membership functions $\mu_{j,i}(x_i^{(p)})$ are calculated using (A.9), and:

$$W_k = -\frac{1}{2} \left(A K_k A^T \right)^{-1} \\ \vec{w}_k = \left(A K_k A^T \right)^{-1} \left(A \vec{\mu}_k - A \vec{\mu} \right) \quad , \text{ for } k=1, 2, \dots, c \quad (37) \\ w_{0,k} = -\frac{1}{2} \left(A \vec{\mu}_k - A \vec{\mu} \right)^T \left(A K_k A^T \right)^{-1} \left(A \vec{\mu}_k - A \vec{\mu} \right) - \frac{1}{2} \ln |A K_k A^T| + \ln P(\omega_k)$$

In the case of equal covariance matrices (A.22), the dependence (36) simplifies slightly to

$$F_k(\vec{x}) = \sum_{\substack{i=1 \\ x_i^{(d)}=df_{j,i}}}^a \ln(p_{j,i}^k) + \sum_{i=1}^t \ln\left(\sum_{j=1}^{h_i} \mu_{j,i}(x_i^{(p)}) q_{j,i}^k\right) + \sum_{i=1}^g p(x_i^{(i)} | \omega_k) + \\ + \vec{w}_k \left(A\vec{x}^{(c)} - A\vec{\mu} \right) + w_{0,k} \quad , \text{ for } k=1, 2, \dots, c \quad (38)$$

where

$$\vec{w}_k = \left(A \left(\sum_{i=1}^c P(\omega_i) K_i \right) A^T \right)^{-1} \left(A \vec{\mu}_k - A \vec{\mu} \right) \quad , \text{ for } k=1, 2, \dots, c \quad (39) \\ w_{0,k} = -\frac{1}{2} \left(A \vec{\mu}_k - A \vec{\mu} \right)^T \left(A \left(\sum_{i=1}^c P(\omega_i) K_i \right) A^T \right)^{-1} \left(A \vec{\mu}_k - A \vec{\mu} \right) + \ln P(\omega_k)$$

A stable numerical scheme to calculate the posterior probabilities from the discriminant functions is

$$P(\omega_k | \vec{x}) = \frac{1}{\sum_{i=1}^c e^{F_i(\vec{x}) - F_k(\vec{x})}}, \text{ for } k=1, 2, \dots, c \quad (40)$$

In general, the classification is done by finding appropriate discriminant functions for each class. Each assigns a real number to the measurement vector. The latter is assigned to that class which has the largest discriminant function value (Klir and Yuan, 1995, p. 358). So, the classification may be conducted using any type of non-linear machine, where the discriminant function may not be interpreted as dependent on the class k part from the natural logarithm of the posterior probability. Then (40) may also be utilized, yet the result would not be some posterior probabilities, but a quasi-probability (Wang and Klir, 2009).

During maintenance of turbomachinery, it is necessary to not only know the status at a given moment, but also judge on the change of that status, its direction and its speed (i.e. how fast does the change occur). Practically speaking, the task here is to develop a fully functional system, which classifies and makes conclusions (forecasts) if there is a fault that might develop in time. If so – then we need to identify that fault and predict when it is going to start to explicitly dominate the technical status of the machine.

Collecting enough dynamic learning information is a complex, expensive, slow and almost impossible process. Therefore, we try to solve the problem of fault prediction based on enough static information to create the classifier. The dynamic information for the process of technical system deterioration is encoded in M count of input observations in various moments of time t_i . This results in a multi-dimensional linear trend with varying critical values determined by the classifier.

Let the forecasting sample X contains M observations $\vec{x}_1, \vec{x}_2, \dots, \vec{x}_M$, measured in times t_1, t_2, \dots, t_M , where $t_i < t_{i+1}$ for $i=1, 2, \dots, M-1$. The purpose is to give one of the following four types of answers:

- 1) There is more than a single most probable class
- 2) The class number γ has tendency to stay the most probable
- 3) The status will change from class γ to class δ
- 4) The status will change from class number γ to class number δ after time $T - t_M$ after the last measurement

The meaning of the first type of answer is to realize the physical essence of forecasting, i.e. the fully-functional status dominates in the analysed moments of time, while faults are only in development, but have not occurred (explicitly) yet. There are many methods to solve the tasks of identifying one of the four types of answers. A novel algorithm is presented below.

Algorithm 1: Prediction without dynamic learning

1. Using hybrid Bayesian classifier, we calculate the discriminant functions $F_k(\vec{x}_i) = F_{k,i}$ for each observation $\vec{x}_1, \vec{x}_2, \dots, \vec{x}_M$, for $i=1, 2, \dots, M$ and each class $k=1, 2, \dots, c$. We need to calculate the most probable class γ in the moment t_1 : $\gamma = \arg\{\max(F_{k,1})\}$. If γ is different from $\arg\{\max(F_{k,1})\}$ for any $i=2, 3, \dots, M$, then the answer of first type is the correct one. We can also assume that γ is the fully-functional class to support the physical essence of forecasting.
2. Recalculate $F_{k,i}$ for each observation $\vec{x}_1, \vec{x}_2, \dots, \vec{x}_M$, for $i=1, 2, \dots, M$ and each class $k=1, 2, \dots, c$ into posterior probabilities $P_{k,i}$ using formula (40).
3. Calculate the backward discriminant functions $G_k(\vec{x}_i) = G_{k,i} = \ln(P_{k,i})$.
4. For each of the k classes, construct the predicted discriminant functions $G_k(t) = a_k t + b_k$, using the pairs $t_1 - G_{k,1}, t_2 - G_{k,2}, \dots, t_M - G_{k,M}$. We can use ordinary least square without weighing.

Alternatively, if we want to increase the influence of the last observations we can use weighted least square with weight coefficients equal to either time t_i (correction by time) or i (correction by observation number)

5. The line $G_\gamma(t)$ crosses the lines $G_k(t)$ for $k \neq \gamma$, where the crossing point has a coordinate T_k . If a_k is lower or equal to a_γ , then T_k is set to $+\infty$.
6. Let T be the minimum of T_k and δ is the argument of the minimum of T_k .
7. If T is $+\infty$ then the second type of answer applies.
8. If $T < t_M$, then the third type of answer applies.
9. If neither of the above is true, then the fourth type of answer applies.

Figure 1 presents an example to demonstrate the fourth type of answer. We have four observations made in four moments of time ($M=4$) – t_1, t_2, t_3 and t_4 to classify the status of an object into $c=3$ classes. We have utilized linear models of type $G_k(t)=a_k t+b_k$. The resulting classification indicates that $\gamma=2$ and $\delta=1$.

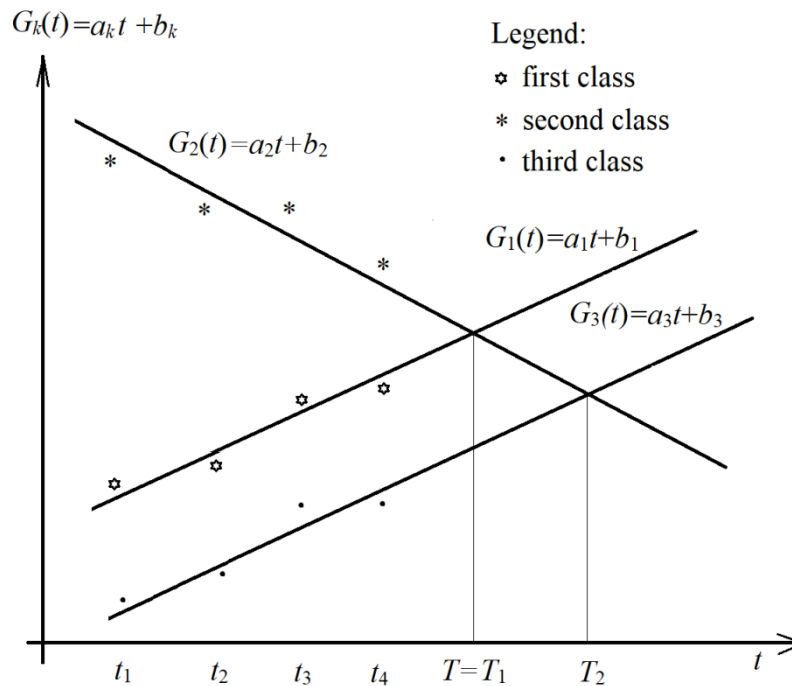


Figure 1. Demonstration of the fourth type of answer for prediction of four measurements into three classes

We selected linear models for each class because of the following considerations:

- If the forecasted effect has started to develop visibly in moment t_j , then there is no point in collecting large sets of observations for the time before t_j
- Defects develop with different speed and it is hard to collect a large set of observations after the moment t_j
- The linear model is largely robust, i.e. insensitive to noise in the input data
- The linear model is applied to various non-linearly related parameters using the discriminant function
- It is not mandatory to have good linear models for every class. Of importance is mostly the quality of the linear models for the classes that define the forecast

5. Practical application

We have applied the proposed hybrid Bayesian Classifier for technical diagnostics of the turbomachines included in an ammonia synthesis loop in a nitrogen fertilizer plant. The technological scheme of the production is given in Figure 2. The feedgas has a discharge of 72000 Nm³/h (normal cubic meters per hour). It contains 73% Hydrogen, 26.2 % Nitrogen, 0.6% Methane, and 0.2% Argonne. Four consecutively connected compressors increase the pressure of the feedgas from 22.5 bar to 270 bars. Anhydrous liquid ammonia with discharge of 2600 kg/m is produced in the ammonia synthesis column. The latter is part of a recirculation loop with a recirculation rate of 5.5 times. The compressors are powered by a SEG 25 steam turbine, which is situated in the middle of the compressor train. The turbine is a single casing, non-reheat two-section, arranged in series, with one inter-section steam extraction. The extracted steam as well as the exhausted steam after the low-pressure section are used for technological purposes. All compressors are centrifugal, single-shaft, multiple stage, barrel type, with vaneless diffusers. The first one, 2BCL408, is an eight stage, with back-to-back arranged two sections, and with inter-stage cooler (H1801). Let us denote the first section with 1A and the second section with 1B. The second compressor (denoted as 2), BCL407 is a seven stage, single section. It sucks the feedgas from the external cooler H1802, which is after the first compressor. The third compressor (denoted as 3), BCL306 is a six stage, single section. It sucks the feedgas from the external cooler H1803, which is after the second compressor. The fourth compressor, 2BCL306, is a 6 stage, with back-to-back arranged two sections. The first section (denoted as 4) of 2BCL306 has five impellers. It sucks the feedgas from the external cooler H1804, which is after BCL306 and discharges the gas in the recirculation loop through the external cooler H1805. The feedgas movement in the recirculation loop is powered by the sixth impeller of 2BCL306. This last stage forms the second section (denoted as 4P) of the compressor.

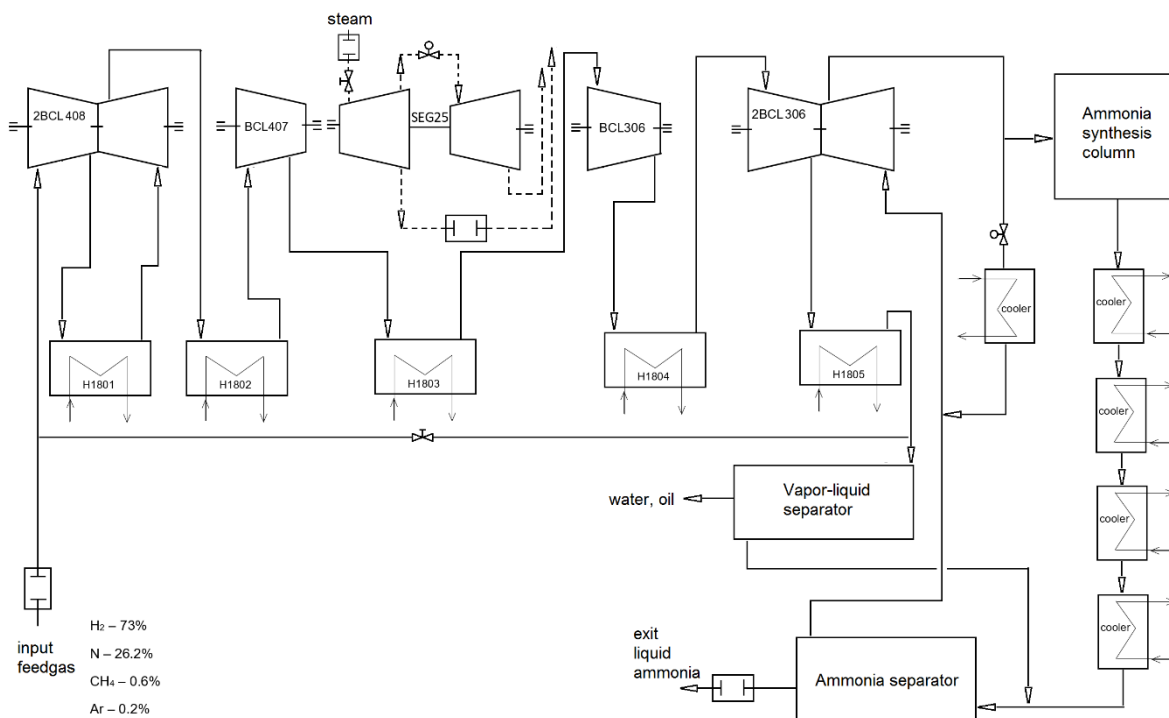


Figure 2. Technological scheme of ammonia production

We have developed series of classifiers for technical diagnostics of different elements of the turbomachinery train. One such classifier is described in (Tenekedjiev et al., 2006). It monitors the gas flow section of the four compressors by determining the posterior probabilities for $c=13$ classes: “no faults” and 12 faults. Six of the faults refer to labyrinth packing aging of the gas side seals. Another six refer to broken rubber O-rings. The observation vector consisted of 34 features: $a=11$

discrete (each with $b_i=3$ nominal value), $t=5$ pseudo-discrete (each with $h_i=2$ typical sub intervals) and $e=18$ multi-normal continuous.

The first 5 discrete features are the presence of oil in the draining cage of each intermediate cooler with $b_1=b_2=b_3=b_4=b_5=3$ nominal values of ‘no oil’, ‘traces of oil’, and ‘obvious oil presence’. The next 5 discrete features represent the type of compression ratio drop in each of the compressor section 1A, 1B, 2, 3, and 4. The nominal values for any of these features are: ‘no obvious decrease of the actual compression ratio from the scheduled one’, ‘smooth decrease of the actual compression ratio from the scheduled one’, and ‘sharp decrease of the actual compression ratio from the scheduled one’. The last discrete feature is the type of compression increment decrease in the recirculation section 4P with nominal values of ‘no obvious decrease of the actual compression increment from the scheduled one’, ‘smooth decrease of the actual compression increment from the scheduled one’, and ‘sharp decrease of the actual compression increment from the scheduled one’.

The 5 pseudo-discrete features were formed as the quotients of the actual compression ratios and the scheduled compressor ratios at the current working regime for 1A, 1B, 2, 3 and 4. For all pseudo-discrete features the first typical sub-interval ‘no increase’ was set from $D_{1,i}=0.5$ to $U_{1,i}=1$, whereas the second typical sub-interval ‘increase’ was set from $D_{2,i}=1.02$ to $U_{2,i}=2$.

The first multi-normal continuous feature was the RPM of the compressors shafts. The next 10 multi-normal features were selected as the input and output temperature (in °C) of the gas mixture for 1A, 1B, 2, 3 and 4. The next 5 multi-normal features represented the discrepancies between the compression ratios and the nominal compression ratios at the current working regime for 1A, 1B, 2, 3 and 4. The 17th multi-normal feature is the discrepancy (in bars) between the actual output pressure and the scheduled output pressure at the current working regime at the exhaust valve of 4. The 18th multi-normal feature is the discrepancy (in bars) between actual compression increment and the scheduled compression increment at the current working regime for the circulation section 4P.

We used both a training sample and expert estimates in the conditional likelihood density parameter estimation (with the expert estimates confidence coefficients being very high for the conditional probabilities of the discrete and pseudo-discrete features and satisfactory for those of the multi-normal features). Learning was performed with identical within-class covariance matrix for all classes, estimated mainly based on the ‘no fault’ training sample information. After several improvements, the demonstrated system achieved 93% of agreement between the expert panel opinion and the system output over artificially generated samples.

Table 1 gives information about four measurements of the system, made on the 15th, 45th, 75th and 135th day after main repair. We perform prediction in the 13 classes previously defined and calculate the discriminant functions $F_{k,i}$ in all four measurement points ($i=1, 2, 3, 4$) (see rows 2, 3, 4, and 5 of Table 1). Using Algorithm 1, we recalculate the discriminant functions $F_{k,i}$ to posterior probabilities $P_{k,i}$ first (see rows 6, 7, 8 and 9 of Table 1), and then to backward discriminant functions $G_{k,I}$ (see rows 10, 11, 12 and 13 of Table 1). For each class, the four backward discriminant functions are then approximated with linear models, whose linear parameters a_k and b_k are given in rows 14 and 15 of Table 1. The result of the prediction is that based on correction by item number, step 9 of Algorithm 1 makes a prediction of type 4 as follows: the most probable is class 8 – “labyrinth packing aging of compressor 3” and the aging will occur up to 81 days after last measurement. The result is also demonstrated in Figure 3 where the two important classes – 1 and 8 – are depicted. Row 16 of Table 1 gives the R^2 of the linear models. For classes 1 and 8 we have high R^2 values (of 0.74 and 0.78 respectively), so the linear models are of good quality. This does not hold for the models of some of the other classes (reaching R^2 as low as 0.013). However, this does not diminish the reliability of prediction because in this case the important classes are 1 and 8. The other two methods (no correction and correction by time) give similar results. For the case of no correction the R^2 values for classes 1 and 8 are respectively 0.72 and 0.73 with $T_k=190$ and the aging will occur up to 54 days after last measurement. For the case of correction by time the R^2 values for classes 1 and 8 are respectively 0.81 and 0.85 with $T_k=230$ and the aging will occur up to

96 days after last measurement. Evidently, predictions are practically the same for the three methods. This demonstrates that the prediction is robust.

Table 1. Two types of discriminant functions, posterior probabilities, coefficients of the linear model, R^2 and T_k of prediction problem with four measurements in 13 classes

classes	1	2	3	4	5	6	7	8	9	10	11	12	13
$F_{k,1}$	-2.5e+0	-7.2e+0	-1.1e+1	-8.1e+0	-8.7e+0	-1.0e+1	-9.4e+0	-6.4e+0	-5.4e+0	-6.2e+0	-1.0e+1	-6.0e+0	-8.3e+0
$F_{k,2}$	-1.3e+0	-5.5e+0	-6.4e+0	-6.3e+0	-4.7e+0	-4.8e+0	-4.8e+0	-3.3e+0	-4.4e+0	-5.5e+0	-4.6e+0	-5.7e+0	-5.2e+0
$F_{k,3}$	-1.1e+0	-5.3e+0	-4.7e+0	-5.8e+0	-4.1e+0	-6.1e+0	-6.1e+0	-3.1e+0	-5.9e+0	-3.6e+0	-4.7e+0	-5.8e+0	-3.6e+0
$F_{k,4}$	-1.2e+1	-1.7e+1	-1.7e+1	-1.7e+1	-1.7e+1	-1.8e+1	-1.7e+1	-1.4e+1	-1.4e+1	-1.6e+1	-1.6e+1	-1.8e+1	-1.7e+1
$P_{k,1}$	+8.7e-1	+7.9e-3	+1.3e-4	+3.0e-3	+1.7e-3	+3.4e-4	+8.2e-4	+1.6e-2	+4.5e-2	+2.0e-2	+3.3e-4	+2.6e-2	+2.6e-3
$P_{k,2}$	+7.3e-1	+1.1e-2	+4.2e-3	+4.6e-3	+2.3e-2	+2.1e-2	+2.2e-2	+9.6e-2	+3.1e-2	+1.1e-2	+2.7e-2	+8.5e-3	+1.5e-2
$P_{k,3}$	+6.9e-1	+1.0e-2	+1.8e-2	+6.2e-3	+3.3e-2	+4.3e-3	+4.7e-3	+9.4e-2	+5.3e-3	+5.7e-2	+1.9e-2	+6.4e-3	+5.3e-2
$P_{k,4}$	+6.7e-1	+5.0e-3	+4.4e-3	+7.7e-3	+7.6e-3	+2.3e-3	+5.4e-3	+1.8e-1	+7.7e-2	+1.7e-2	+1.7e-2	+2.3e-3	+6.0e-3
$G_{k,1}$	-1.3e-1	-4.8e+0	-8.9e+0	-5.8e+0	-6.4e+0	-8.0e+0	-7.1e+0	-4.1e+0	-3.1e+0	-3.9e+0	-8.0e+0	-3.6e+0	-5.9e+0
$G_{k,2}$	-3.2e-1	-4.5e+0	-5.5e+0	-5.4e+0	-3.8e+0	-3.8e+0	-3.8e+0	-2.3e+0	-3.5e+0	-4.5e+0	-3.6e+0	-4.8e+0	-4.2e+0
$G_{k,3}$	-3.7e-1	-4.6e+0	-4.0e+0	-5.1e+0	-3.4e+0	-5.4e+0	-5.4e+0	-2.4e+0	-5.2e+0	-2.9e+0	-4.0e+0	-5.1e+0	-2.9e+0
$G_{k,4}$	-4.0e-1	-5.3e+0	-5.4e+0	-4.9e+0	-4.9e+0	-6.1e+0	-5.2e+0	-1.7e+0	-2.6e+0	-4.1e+0	-4.1e+0	-6.1e+0	-5.1e+0
a_k	-1.5e-3	-6.6e-3	+1.1e-2	+6.4e-3	-2.2e-3	-3.7e-3	+1.4e-3	+1.3e-2	+1.1e-2	-1.2e-3	+1.3e-2	-1.7e-2	-5.1e-3
b_k	-2.2e-1	-4.3e+0	-6.3e+0	-5.7e+0	-4.2e+0	-5.3e+0	-5.3e+0	-3.4e+0	-4.6e+0	-3.7e+0	-5.5e+0	-3.7e+0	-3.9e+0
R^2	+7.4e-1	+7.1e-1	+2.7e-1	+9.1e-1	+7.3e-2	+5.3e-2	+5.9e-2	+7.8e-1	+1.9e-1	+1.3e-2	+3.5e-1	+9.6e-1	+7.4e-2
T_k	$+\infty$	$+\infty$	+4.9e+2	+6.9e+2	$+\infty$	$+\infty$	+1.8e+3	+2.2e+2	+3.4e+2	+1.1e+4	+3.6e+2	$+\infty$	$+\infty$

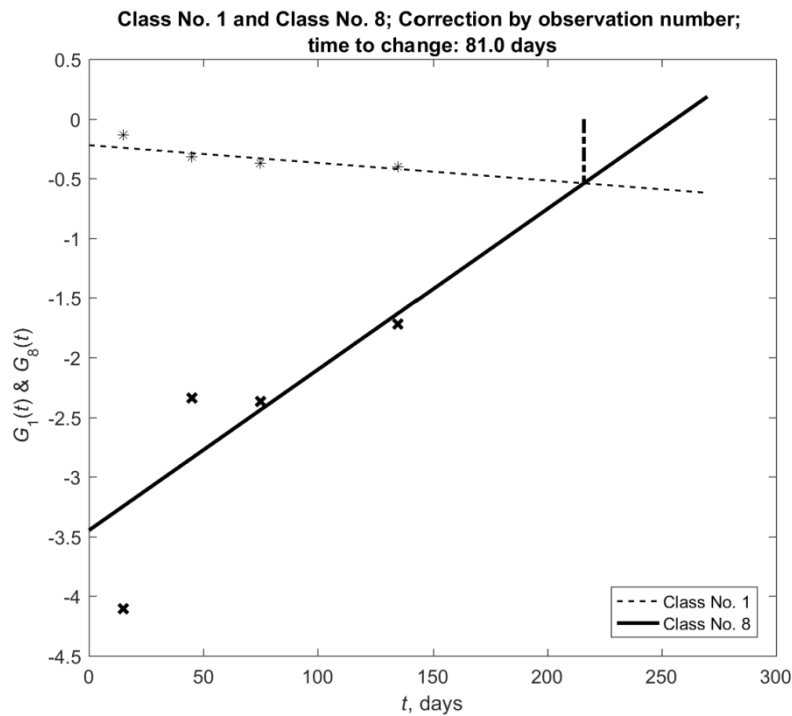


Figure 3. Linear trend of the modified discriminant functions on backward discriminant nodes for classes 8 and 1. The dashed-dotted line is the point of change of the class

In this paper, we describe classifiers of technical state of the journal bearings of the turbomachinery train. The four radial bearings of 2BCL408 and BCL407 are hydrodynamic, split journal bearings with four ball and socket tilting pads each. The six radial bearings of the compressors BCL306, 2BCL306 and the steam turbine SEG25 are hydrodynamic split journal bearings with five self-aligning tilting pads each. A tilting pad is free to rotate about a pivot mounted on the inner housing. When rotating, the shaft is supported by the oil wedges between the sharp and the pads. This improves both the stability of the oil film and the rotor-dynamic stability.

For each bearing, we measure a vibro-signal in the horizontal and vertical plane using piezoelectric sensors, fixed to the bearing housing. The sensor signal is proportional to the vibro-acceleration, which is integrated once to obtain the vibro-velocity. There are two additional eddy-current sensors mounted on each bearing except for the outer bearing of 2BCL408 and 2BCL306. The eight eddy-current sensors measure a signal proportional to the relative vibro-displacement between the shaft and the pad support disk. They are fixed to the pad support disks through the outer housing.

All 36 vibro-signals can be recorded and transferred to a PC for pre-processing. The root mean square (RMS) and the autocorrelation function of each signal is calculated. The RMS is proportional to the vibro-energy. The autocorrelation function is transformed to amplitude spectral density (ASD) using the Fourier integral. The vibro-displacement is fully and automatically compensated for deviations of ideal concentricity of the shaft and for the electrical non-homogeneity of the path of the rotor, on which the sensor is pointed. This is possible thanks to an additional processor, which subtracts a signal measured at low RPM of the shaft from the measured vibro-displacement.

Experimental observations demonstrate that the vibro-signals:

- only slightly depend on the type of bearings'
- depend on the plane, in which the sensors are placed – vertical or horizontal
- strongly depend on the type of sensors – absolute or relative
- strongly depend on the degree of load of the compressor (i.e. on the RPM of the turbine)

As a result, two types of vibro-classifiers are constructed:

- first type that processes the vibro-velocity measured by the two absolute sensors at nominal load of the compressor
- second type that processes vibro-displacement measured by the two relative sensors at nominal load of the compressor

The first type is utilized for diagnostics of the 10 journal bearings of the turbomachinery train. The second type is utilized for diagnostics of the 8 inner journal bearings of the turbomachinery train. Each of the 18 classifiers defines the technical state as one of the following $c=6$ classes: 1) “no faults”; 2) “rotating unbalance of the shaft”; 3) “angular misalignment”; 4) “aging of the pads”; 5) “increased maximum static clearance between the shaft and the pads”; 6) “disrupted lubrication”.

The observation vector consists of 15 continuous features:

1. Percentage change of the actual increase from the input to the output temperature of the bearing lubricant (in % from the scheduled increase at the current working condition from the input to the output temperature of the bearing lubricant).

2. RMS of the vibro-signal on the horizontal plane (in mm/sec for the absolute sensors and in μm for the relative sensors).

3. Amplitude of first harmonic for the horizontal ASD i.e., the value of the horizontal ASD at the fundamental frequency, (in mm/sec for the absolute sensors and in μm for the relative sensors).

4. Amplitude of the second harmonic for the horizontal ASD i.e., the value of the horizontal ASD at twice fundamental frequency (in % from feature 3).

5. Amplitude of third harmonic for the horizontal ASD i.e., the value of the horizontal ASD at trice fundamental frequency (in % from feature 3).

6. Amplitude of fourth harmonic the horizontal ASD i.e., the value of the horizontal ASD at quadruple the fundamental frequency (in % from feature 3).

7. Amplitude of the maximal subharmonic for the horizontal ASD i.e., the maximal value for the horizontal ASD at a frequency less than the fundamental (in % from feature 3).

8. Amplitude of the maximal high interharmonic for the horizontal ASD i.e., the maximal value for the horizontal ASD at a frequency greater than the quadruple fundamental frequency (in % from feature 3).

9. RMS of the vibro-signal on the vertical plane (in mm/sec for the absolute sensors and in μm for the relative sensors).

10. Amplitude of first harmonic for the vertical ASD value, that is the value of the vertical ASD at the fundamental frequency, (in mm/sec for the absolute sensors and in μm for the relative sensors).

11. Amplitude of the second harmonic for the vertical ASD i.e., the value of the vertical ASD at twice fundamental frequency (in % from feature 10).

12. Amplitude of third harmonic for the vertical ASD i.e., the value of the vertical ASD at three fundamental frequency (in % from feature 10).

13. Amplitude of fourth harmonic the vertical ASD i.e., the value of the vertical ASD at quadruple fundamental frequency (in % from feature 10).

14. Amplitude of the maximal subharmonic for the vertical ASD i.e., the maximal value for the vertical ASD at a frequency less than the fundamental (in % from feature 3).

15. Amplitude of the maximal high interharmonic for the vertical ASD i.e., the maximal value for the vertical ASD at a frequency greater than the quadruple fundamental frequency (in % from feature 10).

Initially, it was assumed that all the 15 coordinates were multi-normal ($e=15$) for the two types of classifiers (the relative and the absolute one). For both classifiers, we used almost exclusively expert estimates in the conditional likelihood density parameter estimation (with the expert estimates confidence coefficients being very high for the mean values, satisfactory for standard deviations and low for the correlation coefficients). The training samples were available mostly for class 1 ("no faults"). The learning was performed with identical within-class covariance matrix for all classes. Initially, the relative sensor classifier achieved 71% of agreement between the expert panel opinion and the system output over artificially generated samples. The same percentage for the absolute sensor classifier happened to be 64%. Since the results were not satisfactory, especially for the absolute sensors, we replaced features 7, 8, 14 and 15 from multi-normal to independent continuous features, estimated on 7 quantile indices:

$\alpha_{0.05}, \alpha_{0.2}, \alpha_{0.4}, \alpha_{0.5}, \alpha_{0.6}, \alpha_{0.8}, \alpha_{0.95}$. The classifier with the absolute sensors substantially improved its results and reached 75% agreement. However, the results from the classifier with the relative sensors deteriorated and dropped to 67%. Therefore, finally the 10 classifiers on absolute sensors work with 15 features, of which $e=11$ multi-normal (1-6, 9-13) and $g=4$ independent continuous features (7, 8, 14, 15). The 8 classifiers on relative sensors work with 15 features, all of which are multi normal ($e=15$).

To improve the precision of classification, we modified the hybrid Bayesian PRS presented. The main goal was to incorporate two empirical rules regarding the proxy vibration V , which is the sum of RMS in the horizontal and the vertical plane:

1) if V is substantially lower than the first marginal rate of the sub-failure vibrations (2.8 mm/sec vibro-velocity and 18 μm vibro-displacement), then the classification should indicate class 1 regardless of the values of the other features.

2) if V is around or above the first marginal rate of the sub-failure vibrations, then the classification should not indicate class 1 regardless the values of the other features.

An additional goal of the modification was to produce a system very sensitive to emerging faults especially at the early stages of fault development. The modification was called classification with pseudo priors:

1) Calculate V as sum of the 2nd and 9th feature.

2) Set the parameters PM_1 and PM_2 according to the type of classifier. (Here: $PM_1=1 \mu\text{m}$ and $PM_2=2 \mu\text{m}$ for absolute sensors, whereas $PM_1=10 \text{ mm/sec}$ and $PM_2=12 \text{ mm/sec}$ for relative sensors).

3) Calculate the pseudo-priors as a function of V (see Figure 4):

$$PP(\omega_1) = \begin{cases} 1 - (c-1)s & \text{for } V \leq PM_1 \\ s + (1-cs) \frac{PM_2 - V}{PM_2 - PM_1} & \text{for } PM_1 < V < PM_2 \\ s & \text{for } PM_2 \leq V \end{cases} \quad (41)$$

$$PP(\omega_k) = \begin{cases} s & \text{for } V \leq PM_1 \\ s + \frac{(1-cs)(V - PM_1)}{(c-1)(PM_2 - PM_1)} & \text{for } PM_1 < V < PM_2 \\ (1-s)/(c-1) & \text{for } PM_2 \leq V \end{cases} \quad (42)$$

Formula (42) holds for $k=2, 3, \dots, c$

4) Add $\ln[PP(\omega_k)]$ to the discriminant function of the k -th class $F_k(\vec{x})$.

The general form of the pseudo-priors for $c=6$ classes of state is given on Figure 4.

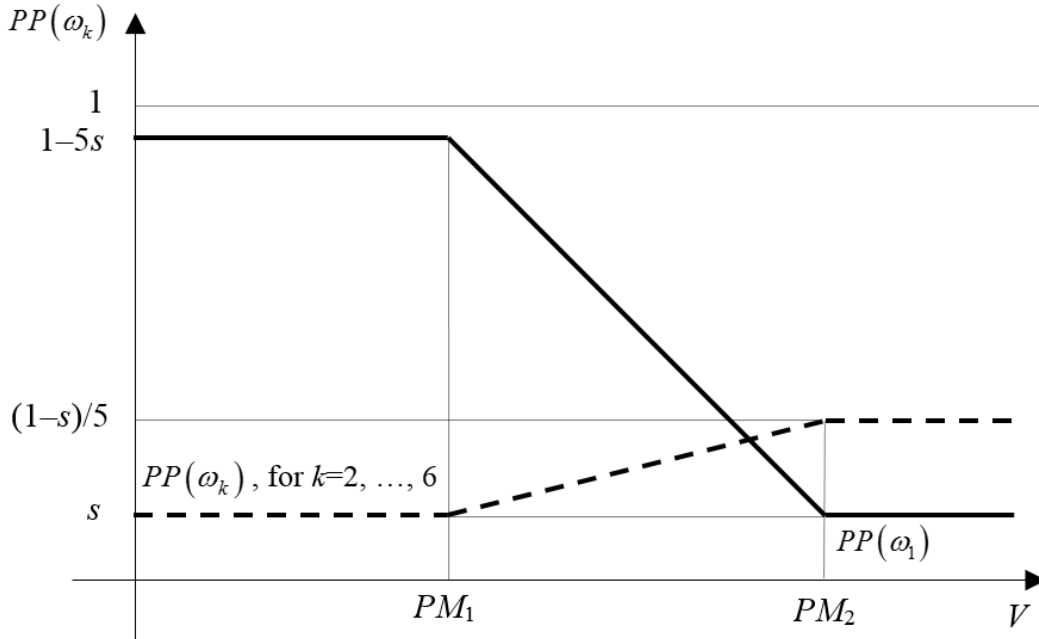


Figure 4. Relation between the proxy vibration V and the pseudo-priors for $c=6$ classes of state

The application of the classification with pseudo priors increased the agreement the agreement between the expert panel opinion and the system output over artificially generated samples for the relative sensor classifier to 80% and to 83% for the one with absolute sensors.

6. Conclusions

Our proposed method for prediction has many advantages, especially when compared to the three widespread alternative approaches – predicting trends of the physical measurement, predicting trends of the posteriors, and predicting trend of the discriminant functions.

The simplest prediction idea is to form linear trends of the coordinates of \vec{x} in the physical space of the observations. However, in the presence of a discrete part $\vec{x}^{(d)}$, it is impossible and incoherent to make solid prediction based on a linear model.

On the other hand, prediction by posterior probabilities is doubtful as in any given moment of time the predicted values $P_1(t), P_2(t), \dots, P_c(t)$ may lose their probabilistic nature (i.e. they may exceed one or be negative or their sum does not add up to one).

A good side of forecasting by discriminant functions $F_k(t)$ is that in any given time it is possible to recalculate them to coherent posterior probabilities. They are defined as that part of the posterior probabilities' logarithm that is dependent on the class. The part that is independent of the class, however, is an arbitrary additive constant. Therefore, from a classification stand point, the arbitrary nature of the constant is of no importance. Yet if we need to capture the trend of change it makes a huge difference (see for example rows 2-5 of Table 1 as well as Figure 1). If we do not know the additive constant, then it is needless to explore the trends as the trends would be purely random. Therefore, forecasting by discriminant functions will give poor results.

It is obvious that the three alternative prediction approaches give questionable results. In our proposed novel algorithm, we use the logarithm of the posterior probabilities to find the backward discriminant functions for nodes of the linear trend. The backward discriminant function is a discriminant function by its own right. However, the predicted discriminant functions $G_k(t)$ are not random because the nodes do not contain an arbitrary additive constant.

The stage of development of the mathematical background of pattern recognition is so high that practically any method would work well if there is a good number of training samples. Unfortunately, such training samples are only present for the no-fault class with training samples for the other classes in very short supply. Such problems are also known as imbalanced and cause huge problems for practical pattern recognition (Barandela et al., 2003). This is another reason why using systems that can utilize subjective information is a recommended approach.

The paper offered a discussion on the ways to represent learning information and the sources for obtaining the later. Three types of parameter estimation of conditional likelihood density functions were incorporated in the hybrid Bayesian classifier: 1) using only training samples; 2) using only expert information; 3) pseudo-Bayesian data fusion approach. These parameters were used to create a non-linear machine for classification consisting of one discriminant function per each class. The final product uses fuzzy concepts in the data presentation, the parameter estimation, the learning and the pattern recognition phases to ensure that the resulting classification is optimum at any quality and quantity of learning information. Such a classifier is of crucial importance for technical diagnostics of equipment in processing industry, where the aggressive production conditions cause rapid aging effects and the learning information increases with the time for operation. We provided example of such an application for turbomachinery in ammonia synthesis.

The hybrid Bayesian PR classifier may have other components in addition to the ones discussed here. It allows integration into larger-scale data processing systems that also perform feature selection, visualization, technical forecasting, clustering, data generation, etc. Those components are beyond the scope of the current paper, yet they speak of the versatile essence of the approach and its practical value.

It is unreasonable to believe that any classification system is applicable for technical diagnostics of any equipment. The heuristic knowledge either from expert opinion or from data mining (Alonso and Espin Andrade, 2013) in any form should be utilized as modifications of the basic nonlinear machine (see Receo-Garcia et al. (2010)). The presented pseudo-priors modification of the hybrid Bayesian PR classifier for the technical diagnostics of the turbomachinery journal bearings is an example of such an approach. This is a rule more than an exception in any practical application (Hadjiski, Doukovska, 2013).

In the hybrid Bayesian PR classifier the training process is brought down to assessment of parameters. There are many methods in QDA about parameter estimation, probability estimation, numerical characteristics estimation (e.g. probability quantiles), etc. Some of those techniques for example would be suitable to construct the conditional distributions of the multi-normal features. If experts apply the PR techniques, then we expect that they have sufficient statistical skills to handle the estimation themselves. For those not so knowledgeable in the field of stochastic uncertainty

assessment there are formalized techniques to extract quantitative estimates of those parameters. This implies that estimates would be subjective, but not arbitrary.

Acknowledgments

This work was supported in part by the Higher Education Institutional Excellence Programme of the Ministry of Human Capacities in Hungary for the thematic programme of Semmelweis University.

References

- Alonso, M.M., Espin Andrade, R.A., 2013. Knowledge discovery by compensatory fuzzy logic predicates using a metaheuristic approach, *Studies in knowledge discovery, knowledge management and decision making*, Fourth international workshop proceedings EUREKA 2013, Mazatlan, November 2013, Mexico, 17-26.
- Bahrapour, S.; Moshiri, B., Salahshour, K., 2011, Weighted and constrained possibilistic C-means clustering for online fault detection and isolation. *Applied Intelligence* 35, 269-284.
- Barandela, R., Sanchez, J.S., Garcia, V., Rangel, E., 2003. Strategies for learning in class imbalance problems. *Pattern Recognition* 36, 849-851.
- Brandan Briones, L., Lazovik, A., Dague, Ph., Optimal observability for diagnosability, *Proc. 19th International Workshop on Principles of Diagnosis*, 22-24 Sept, Australia, 31-38.
- Bregon, A., Biswas, G., Pulido, B., 2008. Compilation techniques for fault detection and isolation: a comparison of three methods, *Proc. 19th International Workshop on Principles of Diagnosis*, 22-24 Sept, Australia, 39-46.
- Brown, R., 2004. *Rational choice and judgment – decision analysis for the decider*, Wiley.
- Cordier, M.-O., Dague, P., Levy, F., Montmain, J., Staroswiecki, M., Trave-Massuyes, L., 2004. Possible conflicts: a compilation technique for consistency-based diagnosis approach from the artificial intelligence and automatic control perspectives, *IEEE Transactions on Systems, Man, and Cybernetics–Part B: Cybernetics* 34(5), 2163– 2177.
- Cullum, J., Tenekedjiev, K., 2017, Implementing risk based maintenance in the naval context: review and directions, *Mechanics of machines XXV(1)*, 60-66.
- Ding, S.X., 2008. *Model-based fault diagnosis techniques*, Springer. 2008
- Ding, Sh., Zhu, H., Jia, W., Su, Ch., 2011. A survey on feature extraction for pattern recognition. *Artificial Intelligence Review* 37(3), 169-180.
- Doraiswami, R., Cheded, L., 2017, Fault detection and isolation, In: *Fault diagnosis and detection*, Demetgul M, Ünal, M. (Eds), 534-580, IntechOpen
- Duda, R., Hart, P., Stork, D., 2001. *Pattern classification*, Second Edition, Wiley.
- Ebeling, Ch., 2010. *An introduction to reliability and maintenance engineering*, Second Edition, Waveland Press Inc.
- Fatichah, C., Tangel, M.L., Yan, F., Betancourt, J.P., Widyanto, M.R., Dong, F., Hirota, K., 2015. Fuzzy feature representation for white blood cell differential counting in acute leukemia diagnosis, *International Journal of Control, Automation and Systems* 13(3), 742-752.
- French, S., 1993. *Decision theory: an introduction to the mathematics of rationality*, Wiley.
- French, S., Insua, D.R., 2010. *Statistical decision theory*, Kendall's Library of Statistics 9, Arnold-UK.
- Gammack, J., Hobbs, V., Pigott, D., 2011. *The book of informatics*, Revised Edition, Cengage.
- Hadden, G., Bergstrom, P., Samad, T., Bennett, B.H., Vachtsevanos, G., van Dyke, J., 2000. System health management for complex systems, In: Samad, T., Weyrauch, J. (Eds). *Automation, control and complexity – an integrated approach*, 191-214.
- Hadjiski, M., Doukoska, L., 2012. Technical diagnostics of mill fan system. *Comptes Rendus de l'Academie Bulgare des Sciences* 65(12), 1731-1738.
- Hadjiski, M., Doukoska, L., 2013. CBR Approach for technical diagnostics of mill fan system. *Comptes Rendus de l'Academie Bulgare des Sciences* 66(1), 93-100.

- Hambebo, B.M., Carvalho, M., Ham, F., A scalable approach to network traffic classification for computer network defense using parallel neural network classifier architecture, In Igelnik, B., Zurada, J. (Eds.) *Efficiency and scalability methods for computational intellect*, 181-196.
- Heng, A., Zhang, S., Tan, A.C.C., Mathew, J., 2009. Rotating machinery prognostics: state of the art, challenges and opportunities, *Mechanical systems and signal processing* 23(3), 724–739.
- Hoti, F., Holmstrom, L., 2003. A Semi-parametric approach to statistical pattern recognition. *Bulletin of the International Statistical Institute* LX(1), 510-511.
- Iserman R., 2006. *Fault-diagnosis systems*, Springer.
- Jain, A.K., Duin, R.P. Mao, J., 2000. Statistical pattern recognition: a review. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 22, 4-37.
- Klir, G., 2006. *Uncertainty and information – foundations of generalized information theory*, Wiley.
- Klir, G., Yuan, B., 1995. *Fuzzy sets and fuzzy logic – theory and applications*, Prentice Hall.
- Krysander, M., Nyberg, M., 2008. Statistical properties and design criteria for fault isolation in noisy systems, *Proc. 19th International Workshop on Principles of Diagnosis*, 22-24 Sept, Australia, 101-108.
- Marinov, B., Kazakov, A., Raykov, P., 2014. Modelling of big engines dynamic behaviour for marine vessels, *Comptes Rendus de l'Academie Bulgare des Sciences* 67(3), 419-426
- Marinov, B., Kazakov, A., Tsveov, M., 2016. Frequency and modal analysis of a five mass power transmission line of a ship, *Comptes Rendus de l'Academie Bulgare des Sciences*, Volume 69(1), 75-84.
- Mosterman, P., Biswas, G., 1999. Diagnosis of continuous valued systems in transient operating regions, *IEEE T. on Systems, Man, and Cybernetics* 29(6), 554–565.
- Nikolova, N.D., Kirilova, E., Georgieva, Ts., Daskalov, P., Tenekedjiev, K., 2013. LABVIEW based virtual instrument for recognition of fusarium moniliforme damaged corn seeds by color image analysis, *Proc. International Conference "Food Science, Engineering, and Technologies 2013"*, 18-19 October, Plovdiv, Bulgaria, 69-74.
- Nikolova, N.D., Makedonska, D., Tenekedjiev, K., 2004. Application of quantile approximated distributions for pattern recognition, described with continuous independent features, *Machine Mechanics*, Section “Diagnostics” 57(3), 103-106.
- Nikolova, N.D., Toneva-Zheynova, D., Stoyanov, B., Tenekedjiev, K., 2010. Models for estimation of correlation coefficients using conditional quartiles for two-dimensional normal distributions. *Entrepreneurship and Innovation Journal* 2(2), 83-94.
- Nilsen, S., 2006. Chapter 12: Challenges to safety management when incorporating integrated operations solutions in the oil industry, In: Svenson, O., Salo, I., Oedewald, P., Reiman, T., Skjerve, A.B. (Eds.) *Nordic perspectives on safety management in high reliability organizations – theory and applications*, 171-182, Akademitryck, Valdemarsvik – Sweden.
- Palade, V., Patton, R.J., Uppal, F.J, Quevedo, J., Daley, S., 2002. Fault diagnosis of an industrial gas turbine using neuro-fuzzy methods, In: *IFAC proceedings Volumes* 35(1), 471-476.
- Patton, R. J., Frank, P. M., Clark, R. N., 2000. *Issues in fault diagnosis for dynamic systems*, Springer Verlag, New York.
- Pedrycz, W., 1997. Fuzzy sets in pattern recognition: accomplishments and challenges. *Fuzzy Sets and Systems* 90, 171-176.
- Pratt, J., Raiffa, H., Schlaifer, R., 2008. *Introduction to statistical decision theory*, MIT Press, Paperback.
- Press, W., Teukolski, S., Vetterling, W.T., Flannery, B., 2007. *Numerical recipes: the art of scientific computing*, Third Edition, Cambridge University Press.
- Pulido, B., Alonso-Gonzalez, C., 2004, Possible conflicts: a compilation technique for consistency-based diagnosis, *IEEE Trans. on Systems, Man, and Cybernetics. Part B: Cybernetics* 34(5), 2192–2206.
- Receo-Garcia, J. A., Diaz-Agudo, B., Sanches-Ruiz, A. A., Gonzales-Calero, P. A., 2010. Lessons learnt in the development of a CBR framework. *Expert Update* 10(1), 9-16.

Samantaray, A.K., Medjaher, K., Bouamama, B.O., Staroswiecki, M., Dauphin- Tanguy, G., 2006. Diagnostic bond graphs for online fault detection and isolation, *Simulation Modelling Practice and Theory* 14, 237–262

Tenekedjiev, K., Kobashikawa, C., Nikolova, N., Hirota, K., 2006. Generic database for hybrid bayesian pattern recognition. *Journal of Advanced Computational Intelligence and Intelligent Informatics* 10(3), 419-431.

The Mathworks, 2018. Matlab Function Reference, The MathWorks, Inc.

Tonon, F., Bernardini, A., Elishakoff, I., 1999. Hybrid analysis of uncertainty: probability, fuzziness and convexity. In: Schueller, G.I., Kafka, P (Eds.). *Safety and Reliability*. Volume 2, 1109-1114.

Vachtsevanos G., Lewis, F., Roemer, M., Hess, A., Wu, B., 2006. *Intelligent fault diagnosis and prognosis for engineering systems*, Hoboken, New Jersey, Wiley & Sons.

Wang, Z., Klir, G., 2009. *Generalized measure theory*, Springer.

Zadeh, L.A., 1978. Fuzzy ets as a basis for a theory of possibility. *Fuzzy Sets and Systems* 1, 3-28.

Zadeh, L.A., 1977. Fuzzy sets and their application to pattern classification and clustering analysis. In: Van Ryzin, J. (Ed.) *Classification and clustering*. Academic Press, NY, 251-299.

Zavoichinskaya, E.B., Zavoychinskii, B.I., 2006. Estimation of structure durability using the state diagnostic and safety requirements. *Engineering and Automation Problems* 5(1), 70-77.

Appendix: Hybrid Bayesian PR System

A1. Hybrid Bayesian PR System Features

Let's assume that the technical state S of an equipment at a given moment of time belongs to one of c classes of state ω_k , for $k=1,2,\dots,c$. A formal description of S often comes in the form of a multi-dimensional observation vector \vec{x} , whose features are measurable. The objective of the Bayesian PR is to assess the c posterior probabilities $P(\omega_k|\vec{x})$ for S to be in the different classes:

$$P(\omega_k|\vec{x}) = \frac{P(\omega_k)f_k(\vec{x}|\omega_k)}{\sum_{i=1}^c P(\omega_i)f_i(\vec{x}|\omega_i)}, \text{ for } k=1, 2, \dots, c \quad (\text{A.1})$$

In the so called Bayes theorem (Bayes, 1763), the densities of the conditional likelihoods $f_k(\vec{x}|\omega_k)$ are proportional to the probability for observing \vec{x} , if S belongs to ω_k . The base rates $P(\omega_k)$ (a.k.a. priors) are the probability that S belongs to ω_k when \vec{x} is not known. They can be assessed subjectively, but better to calculate them proportional to the failure rates for each class (Ebeling, 2010, pp. 28-30).

The observation vector \vec{x} , may contain four types of features:

A1.1) A set of a discrete features df_i , each with b_i possible nominal values $dfd_{1,i}, dfd_{2,i}, \dots, dfd_{b_i,i}$, and grouped into an a -dimensional vector $\vec{x}^{(d)} = (x_1^{(d)}, x_2^{(d)}, \dots, x_a^{(d)})^T$, where $x_i^{(d)}$ is the observed value of df_i ;

A1.2) A set of t pseudo-discrete features pf_i each with h_i typical sub-intervals and grouped into a t -dimensional vector $\vec{x}^{(p)} = (x_1^{(p)}, x_2^{(p)}, \dots, x_t^{(p)})^T$, where $x_i^{(p)}$ is the observed value of pf_i . It is said that $x_i^{(p)}$ has fallen into the j^{th} sub-interval when $x_i^{(p)} \in [U_{j-1,i}, D_{j+1,i}]$, but $x_i^{(p)}$ has fallen into the j^{th} typical sub-interval when $x_i^{(p)} \in [D_{j,i}, U_{j,i}]$, where $D_{j,i}$ and $U_{j,i}$ are respectively the lower and upper margin of the j^{th} typical sub-interval of pf_i :

$$D_{1,i} < U_{1,i} < D_{2,i} < U_{2,i} < \dots < D_{h_i,i} < U_{h_i,i} \quad (\text{A.2})$$

The value of $x_i^{(p)}$ may be any real one;

A1.3) A set of e continuous multi-normal features cf_i , grouped in an e -dimensional vector

$$\vec{x}^{(c)} = (x_1^{(c)}, x_2^{(c)}, \dots, x_e^{(c)})^T, \text{ where } x_i^{(c)} \text{ is the observed value of } cf_i;$$

A1.4) A set of g independent-continuous features if_i , grouped in a g -dimensional vector

$$\vec{x}^{(i)} = (x_1^{(i)}, x_2^{(i)}, \dots, x_g^{(i)})^T, \text{ where } x_i^{(i)} \text{ is the observed value of } if_i.$$

The value of $x_i^{(i)}$ will be in the interval $[z_{0,i}, z_{1,i}]$;

As a result, the observation vector $\vec{x} = (\vec{x}^{(d)T}, \vec{x}^{(p)T}, \vec{x}^{(c)T}, \vec{x}^{(i)T})^T$ is an $(a+t+e+g)$ -dimensional column vector with mixed features.

A2. Conditional Likelihood Density Evaluation

In line with traditions in statistical PR described by Fukunaga (1990) and Duda et al. (2001) it is accepted that the discrete, the pseudo-discrete, and the independent-continuous features are independent of each other and from the multi-normal features for every class of state:

$$f_k(\vec{x} | \omega_k) = P_k(\vec{x}^{(d)} | \omega_k) Q_k(\vec{x}^{(p)} | \omega_k) \varphi_k(\vec{x}^{(c)} | \omega_k) \psi_k(\vec{x}^{(i)} | \omega_k), \text{ for } k=1,2,\dots,c \quad (\text{A.3})$$

In (A.3) $P_k(\vec{x}^{(d)} | \omega_k)$ and $Q_k(\vec{x}^{(p)} | \omega_k)$ are the conditional probabilities that the discrete and the pseudo-discrete features of S take values $\vec{x}^{(d)}$ and $\vec{x}^{(p)}$ whereas $\varphi_k(\vec{x}^{(c)} | \omega_k)$ and $\psi_k(\vec{x}^{(i)} | \omega_k)$ are the conditional likelihood densities at points $\vec{x}^{(c)}$ and $\vec{x}^{(i)}$, if S is in class ω_k .

A2.1. Conditional likelihood density for the discrete features

From the independence of the discrete features it follows that

$$P_k(\vec{x}^{(d)} | \omega_k) = \prod_{i=1}^a P(x_i^{(d)} | \omega_k), \text{ for } k=1,2,\dots,c \quad (\text{A.4})$$

In (A.4), $P(x_i^{(d)} | \omega_k)$ is the probability that the discrete feature df_i takes value $x_i^{(d)}$ if S is in ω_k . Let

$$P(x_i^{(d)} = df_{j,i} | \omega_k) = p_{j,i}^k, \text{ for } i=1,2,\dots,a; j=1,2,\dots,b_i; k=1,2,\dots,c \quad (\text{A.5})$$

This means that $p_{j,i}^k$ is the probability of $x_i^{(d)}$ taking the nominal value $df_{j,i}$, if S is in ω_k .

So,

$$P_k(\vec{x}^{(d)} | \omega_k) = \prod_{\substack{i=1 \\ x_i^{(d)} = df_{j,i}}}^a p_{j,i}^k, \text{ for } k=1,2,\dots,c \quad (\text{A.6})$$

A2.2. Conditional likelihood density for the pseudo-discrete features

From the independence of the pseudo-discrete features it follows that

$$Q_k(\vec{x}^{(p)} | \omega_k) = \prod_{i=1}^t P(x_i^{(p)} | \omega_k), \text{ for } k=1,2,\dots,c \quad (\text{A.7})$$

In (A.7), $P(x_i^{(p)} | \omega_k)$ is the probability that pseudo-discrete feature pf_i takes value $x_i^{(p)}$, if S is in ω_k . Let

$$P(x_i^{(p)} \in [D_{j,i}, U_{j,i}] | \omega_k) = q_{j,i}^k, \text{ for } i=1,2,\dots,t; j=1,2,\dots,h_i; k=1,2,\dots,c \quad (\text{A.8})$$

The parameter $q_{j,i}^k$ is the probability of $x_i^{(p)}$ falling into the j^{th} typical sub-interval of pf_i , if S is in ω_k , provided that $x_i^{(p)}$ belongs to one of the typical sub-intervals. The probability $P(x_i^{(p)} | \omega_k)$ is calculated as a function of all the $q_{j,i}^k$ (for $j=1,2,\dots, h_i$) with the help of h_i fuzzy membership functions $\mu_{j,i} = \mu_{j,i}(x_i^{(p)})$ for each of the feature pf_i typical sub-intervals. An example for the seventh pseudo-discrete coordinate $x_7^{(p)}$ with four typical subintervals ($h_7=4$) is given in Figure A.1. The j^{th} fuzzy membership function is 1 for any typical value in $[D_{j,i}; U_{j,i}]$, but linearly decreases to 0 in $U_{j-1,i}$ and in $D_{j+1,i}$:

$$\mu_{j,i}(x_i^{(p)}) = \begin{cases} 1 & \text{if } x_i^{(p)} \leq D_{1,i} \text{ and } j=1 \\ 0 & \text{if } x_i^{(p)} \leq U_{j-1,i} \text{ and } j>1 \\ \frac{x_i^{(p)} - U_{j-1,i}}{D_{j,i} - U_{j-1,i}} & \text{if } x_i^{(p)} \in (U_{j-1,i}, D_{j,i}), \text{ and } j>1 \\ 1 & \text{if } x_i^{(p)} \in [D_{j,i}, U_{j,i}] \\ \frac{D_{j+1,i} - x_i^{(p)}}{D_{j+1,i} - U_{j,i}} & \text{if } x_i^{(p)} \in (U_{j,i}, D_{j+1,i}), \text{ and } j<h_i \\ 0 & \text{if } x_i^{(p)} \geq D_{j+1,i} \text{ and } j<h_i \\ 1 & \text{if } x_i^{(p)} \geq U_{h_i,i} \text{ and } j=h_i \end{cases} \quad (\text{A.9})$$

From (A.9) it is obvious that for each value of $x_i^{(p)}$ the sum of all membership functions will be 1. Then,

$$P(x_i^{(p)} | \omega_k) = \sum_{j=1}^{h_i} \mu_{j,i}(x_i^{(p)}) q_{j,i}^k, \quad \text{for } i=1,2,\dots, t; k=1,2,\dots, c \quad (\text{A.10})$$

From (A.7) and (A.10) follows that,

$$Q_k(\bar{x}^{(p)} | \omega_k) = \prod_{i=1}^t \sum_{j=1}^{h_i} \mu_{j,i}(x_i^{(p)}) q_{j,i}^k, \quad \text{for } k=1,2,\dots, c \quad (\text{A.11})$$

The theoretical justification for applying the probability description of fuzzy measurements (in the form of pseudo-discrete features) is given by Tenekedjiev et al., (2006).

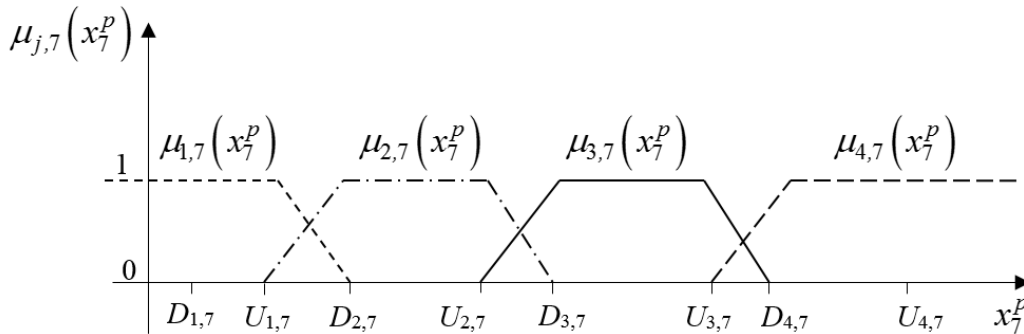


Figure A.1. Fuzzy membership functions for the seventh pseudo-discrete coordinate $x_7^{(p)}$ with four typical subintervals ($h_7=4$)

A2.3. Conditional likelihood density for the multi-normal features

The conditional likelihood density function of $\vec{x}^{(c)}$ is multi-normally distributed, if S is in ω_k . This density is fully characterized with the vector of the mean values $\vec{\mu}_k$ and with the covariance matrix K_k . The vector of the mean values $\vec{\mu}_k$ is an e -dimensional column vector with i^{th} coordinate μ_i^k being the conditional mean value of the continuous multi-normally distributed feature cf_i if S is in ω_k :

$$\vec{\mu}_k = (\mu_1^k, \mu_2^k, \dots, \mu_e^k)^T, \text{ for } k=1, 2, \dots, c \quad (\text{A.12})$$

The covariance matrix K_k is a symmetric positively semi-definite square matrix e by e . Its $(i,j)^{\text{th}}$ entry is $r_{i,j}^k \sigma_i^k \sigma_j^k$, where σ_i^k , σ_j^k , and $r_{i,j}^k$ are the standard deviations and the correlation coefficient of the continuous multi-normally distributed features cf_i and cf_j , if S is in ω_k :

$$K_k = \begin{pmatrix} (\sigma_1^k)^2 & r_{1,2}^k \sigma_1^k \sigma_2^k & \dots & r_{1,e}^k \sigma_1^k \sigma_e^k \\ r_{2,1}^k \sigma_2^k \sigma_1^k & (\sigma_2^k)^2 & \dots & r_{2,e}^k \sigma_2^k \sigma_e^k \\ \vdots & \vdots & \ddots & \vdots \\ r_{e,1}^k \sigma_e^k \sigma_1^k & r_{e,2}^k \sigma_e^k \sigma_2^k & \dots & (\sigma_e^k)^2 \end{pmatrix}, \text{ for } k=1, 2, \dots, c \quad (\text{A.13})$$

The unconditional likelihood density function of $\vec{x}^{(c)}$ is partially characterized by its vector of the mean values $\vec{\mu}$ and its covariance matrix K :

$$\vec{\mu} = \sum_{k=1}^c P(\omega_k) \vec{\mu}_k \quad (\text{A.14})$$

$$K = \sum_{k=1}^c P(\omega_k) K_k + \sum_{k=1}^c P(\omega_k) (\vec{\mu}_k - \vec{\mu})(\vec{\mu}_k - \vec{\mu})^T \quad (\text{A.15})$$

The first sum in (A.15) is the within-class covariance matrix K_W , whereas the second sum in (A.15) is the between-class covariance matrix K_B . The matrix K turns out to be a symmetric positively semi-definite matrix because of (A.15) and because all K_k are symmetric positively semi-definite matrices. So K has e orthonormal eigenvectors \vec{V}_k with associated eigenvalues λ_k . (see Press et al. (2007)). The latter are real, non-negative. We can sort in descending order the eigenvalues and their eigenvectors, so that only the first d are positive:

$$\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_d > \lambda_{d+1} = \lambda_{d+2} = \dots = \lambda_e = 0 \quad (\text{A.16})$$

Let us define a $[d \times e]$ matrix A with i^{th} row containing the transposed i^{th} eigenvector divided to the square root of its eigenvalue:

$$A = \begin{pmatrix} \vec{V}_1^T / \sqrt{\lambda_1} \\ \vec{V}_2^T / \sqrt{\lambda_2} \\ \vdots \\ \vec{V}_d^T / \sqrt{\lambda_d} \end{pmatrix} \quad (\text{A.17})$$

Then, $\vec{x}^{(c)}$ can be linearly transformed from the original e -dimensional space with the multi-normal features cf_i (called physical space) into a d -dimensional space with the continuous multi-normally distributed features cf_{oi} (called objective space):

$$\vec{y}^{(c)} = A\vec{x}^{(c)} - A\vec{\mu} \quad (\text{A.18})$$

The density function of $\vec{y}^{(c)}$ is characterized by a zero vector of the mean values $\vec{\mu}_y$ and identity covariance matrix K_y . So, the objective space is a non-singular, normalized, and centred space which contains the same information as the physical space. The transformation (A.18) is a whitening Karhunen-Loeve transformation (Ding et al., 2011). It is easy for humans to observe the

objects in the objective space by selection two or three coordinates at a time. However, the main advantage is that the covariance matrices in the objective space are (or can be approximated with) invertible ones.

Since the conditional likelihood density function of $\vec{y}^{(c)}$, provided that S is in ω_k , is multi-normally distributed, then it can be calculated using the vector of the mean values $\vec{\mu}_{y,k}$ and the covariance matrix $K_{y,k}$:

$$\vec{\mu}_{y,k} = A\vec{\mu}_k - A\vec{\mu}, \text{ for } k=1,2,\dots,c \quad (\text{A.19})$$

$$K_{y,k} = AK_kA^T, \text{ for } k=1,2,\dots,c \quad (\text{A.20})$$

$$\varphi_k(\vec{x}^{(c)} | \omega_k) = \varphi_k(\vec{y}^{(c)} | \omega_k) = \frac{\exp\left[-\frac{1}{2}\left(\vec{y}^{(c)} - \vec{\mu}_{y,k}\right)^T K_{y,k}^{-1} \left(\vec{y}^{(c)} - \vec{\mu}_{y,k}\right)\right]}{(2\pi)^{\frac{d}{2}} |K_{y,k}|^{\frac{1}{2}}}, \text{ for } k=1,2,\dots,c \quad (\text{A.21})$$

In (A.21), $|K_{y,k}|$ stands for the determinant of $K_{y,k}$, and $K_{y,k}^{-1}$ stands for the inverse of $K_{y,k}$. It is obvious that the covariance matrices $K_{y,k}$ should be all be positively definite in order to be invertible. That is often true in the objective space, but if that is not the case then some approximation should be used. The standard assumption under the circumstances is that the class covariance matrices $K_{y,k}$ are equal and should be substituted by the within-class covariance matrix in the objective space which is non-singular:

$$K_{y,1} \approx K_{y,2} \approx \dots \approx K_{y,d} \approx A\left(\sum_{k=1}^c P(\omega_k)K_k\right)A^T \quad (\text{A.22})$$

A2.4. Conditional likelihood density for the independent continuous features

From the independence of the independent-continuous features it follows that

$$\psi_k(\vec{x}^{(i)}) = \prod_{i=1}^g p\left(x_i^{(i)} | \omega_k\right), \text{ for } k=1,2,\dots,c. \quad (\text{A.23})$$

In (A.23), $p\left(x_i^{(i)} | \omega_k\right)$ is the conditional likelihood density function of the independent-continuous feature if_i , if S is in ω_k . This type of features is discussed by Nikolova et al. (2004). Let us select m quantile names, ‘well’ distributed in the unit interval $[0; 1]$:

$$0 < \alpha_1 < \alpha_2 < \dots < \alpha_m < 1 \quad (\text{A.24})$$

Let

$$\int_{-\infty}^{z_{\alpha_j,i}^k} p\left(x_i^{(i)} | \omega_k\right) dx_i^{(i)} = \alpha_j, \text{ for } i=1,2,\dots,g; j=1,2,\dots,m; k=1,2,\dots,c \quad (\text{A.25})$$

In (A.25) $z_{\alpha_j,i}^k$ is the α_j -quantile from the conditional distribution of the independent-continuous feature $x_i^{(i)}$, if S is in ω_k . We can set up:

$$\alpha_0 = 0 ; z_{0,i}^k = z_{0,i} ; \alpha_{m+1} = 1 ; z_{1,i}^k = z_{1,i} \quad (\text{A.26})$$

Then, the conditional likelihood density of if_i can be approximated as:

$$\begin{aligned}
p\left(x_i^{(i)} \mid \omega_k\right) = & \\
= & \begin{cases} \frac{\alpha_1}{z_{\alpha_1,i}^k - z_{0,i}^k} & \text{if } x_i^{(i)} = z_{0,i}^k \\ \frac{\alpha_{j+1} - \alpha_j}{z_{\alpha_{j+1},i}^k - z_{\alpha_j,i}^k} & \text{if } x_i^{(i)} \in \left(z_{\alpha_j,i}^k, z_{\alpha_{j+1},i}^k\right), \text{ for } j = 0, 1, 2, \dots, m \text{ for } i=1, 2, \dots, g; k=1, 2, \dots, c \text{ (A.27)} \\ \frac{\alpha_{j+1} - \alpha_{j-1}}{z_{\alpha_{j+1},i}^k - z_{\alpha_{j-1},i}^k} & \text{if } x_i^{(i)} = z_{\alpha_j,i}^k, \text{ for } j = 1, 2, \dots, m \\ \frac{1 - \alpha_m}{z_{1,i}^k - z_{m,i}^k} & \text{if } x_i^{(i)} = z_{1,i}^k \end{cases}
\end{aligned}$$

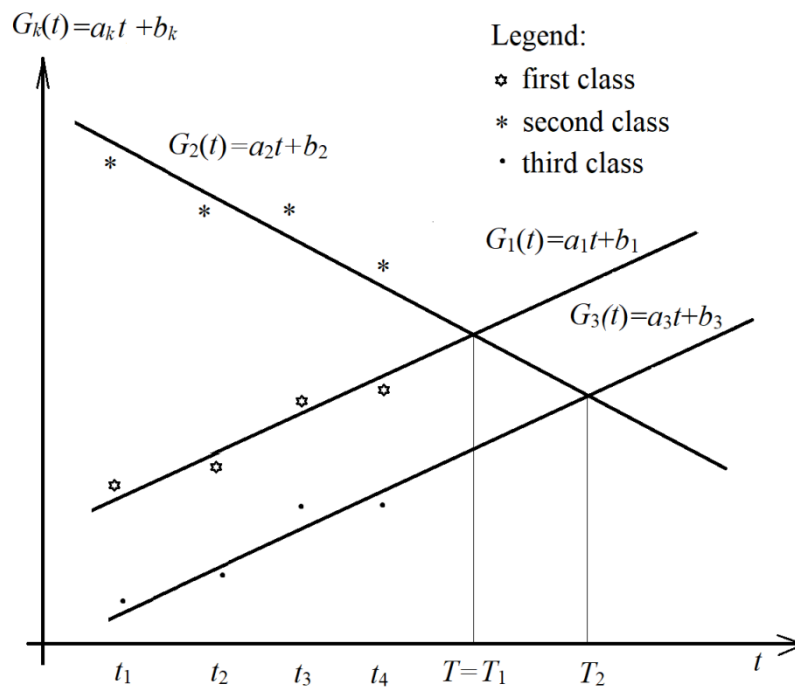


Figure 1. Demonstration of the fourth type of answer for prediction of four measurements into three classes

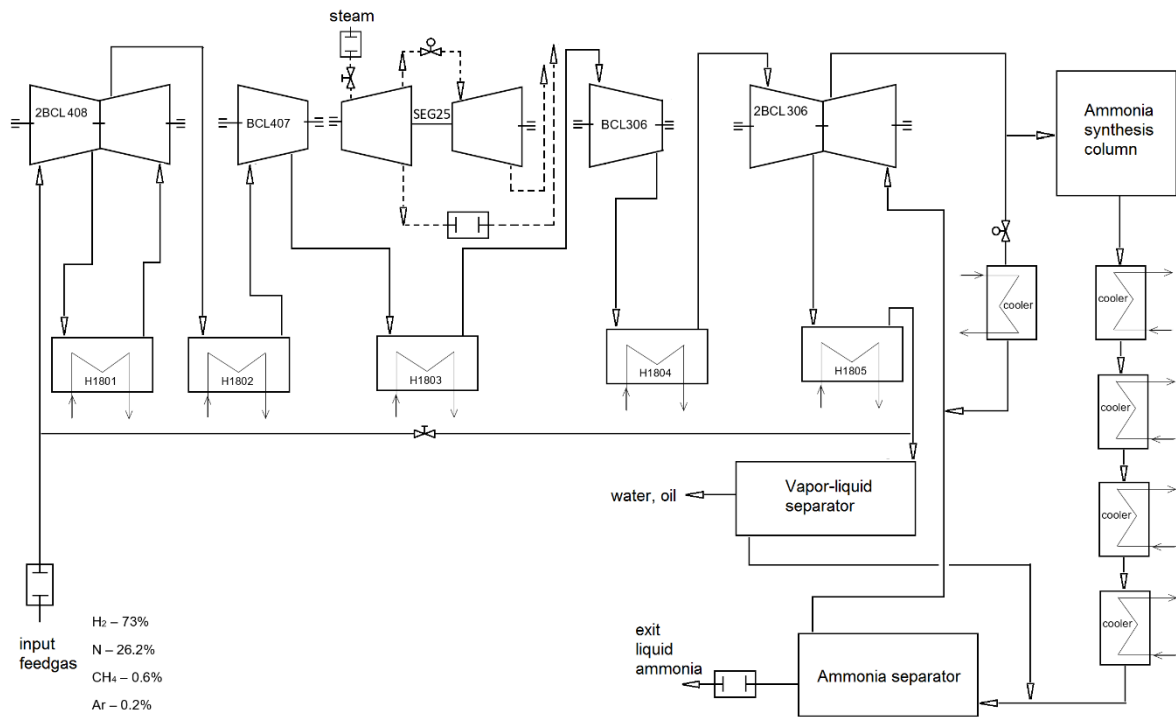


Figure 2. Technological scheme of ammonia production

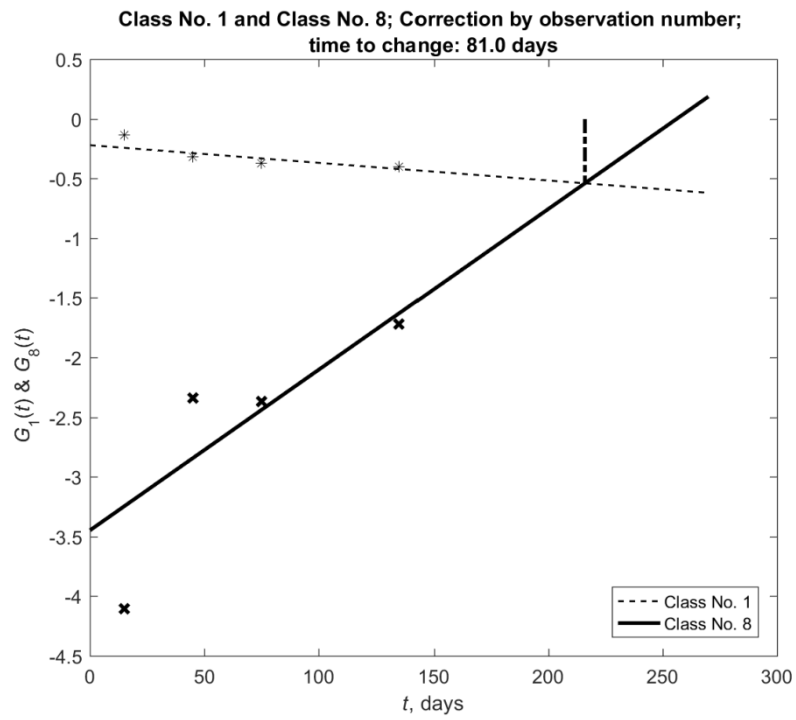


Figure 3. Linear trend of the modified discriminant functions on backward discriminant nodes for classes 8 and 1. The dashed-dotted line is the point of change of the class

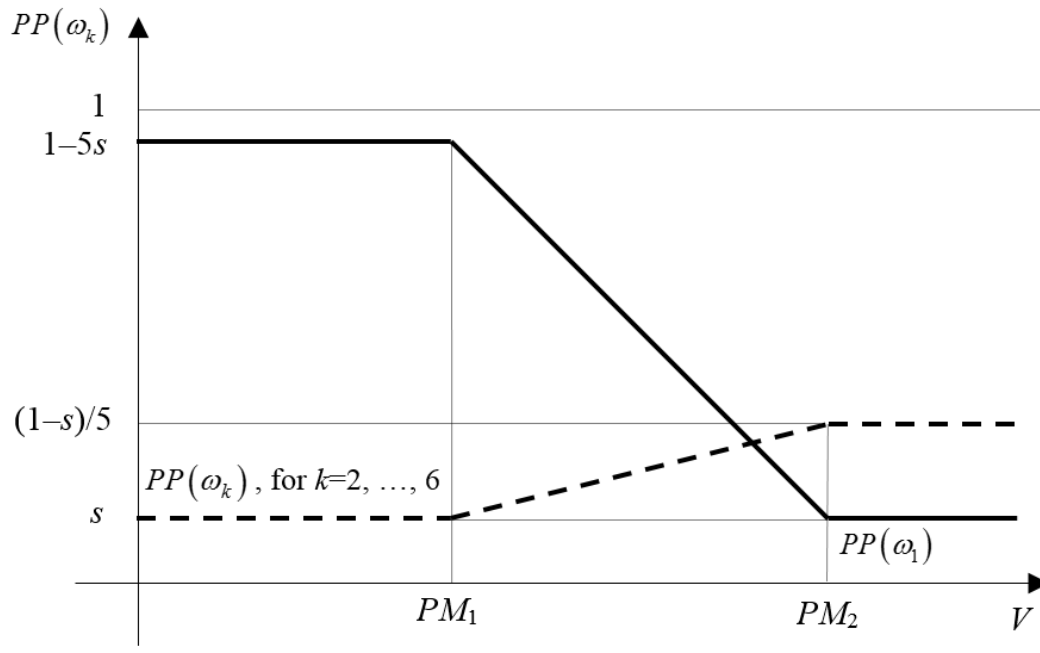


Figure 4. Relation between the proxy vibration V and the pseudo-priors for $c=6$ classes of state

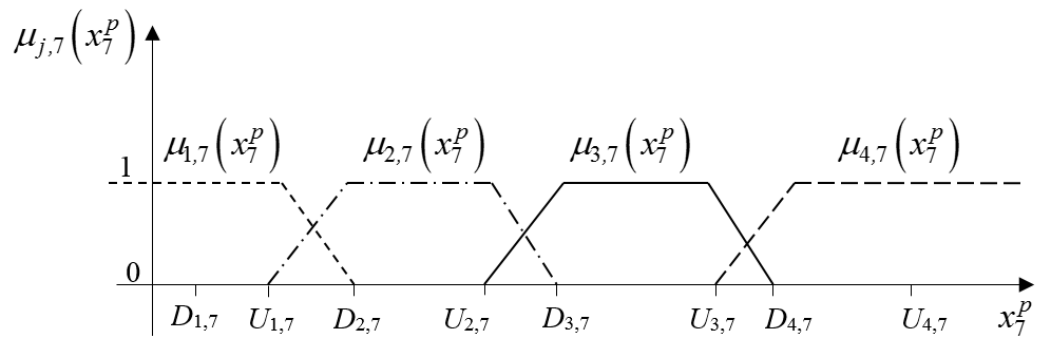


Figure A.1. Fuzzy membership functions for the seventh pseudo-discrete coordinate $x_7^{(p)}$ with four typical subintervals ($h_7=4$)

Table 1. Two types of discriminant functions, posterior probabilities, coefficients of the linear model, R^2 and T_k of prediction problem with four measurements in 13 classes

classes	1	2	3	4	5	6	7	8	9	10	11	12	13
$F_{k,1}$	-2.5e+0	-7.2e+0	-1.1e+1	-8.1e+0	-8.7e+0	-1.0e+1	-9.4e+0	-6.4e+0	-5.4e+0	-6.2e+0	-1.0e+1	-6.0e+0	-8.3e+0
$F_{k,2}$	-1.3e+0	-5.5e+0	-6.4e+0	-6.3e+0	-4.7e+0	-4.8e+0	-4.8e+0	-3.3e+0	-4.4e+0	-5.5e+0	-4.6e+0	-5.7e+0	-5.2e+0
$F_{k,3}$	-1.1e+0	-5.3e+0	-4.7e+0	-5.8e+0	-4.1e+0	-6.1e+0	-6.1e+0	-3.1e+0	-5.9e+0	-3.6e+0	-4.7e+0	-5.8e+0	-3.6e+0
$F_{k,4}$	-1.2e+1	-1.7e+1	-1.7e+1	-1.7e+1	-1.7e+1	-1.8e+1	-1.7e+1	-1.4e+1	-1.4e+1	-1.6e+1	-1.6e+1	-1.8e+1	-1.7e+1
$P_{k,1}$	+8.7e-1	+7.9e-3	+1.3e-4	+3.0e-3	+1.7e-3	+3.4e-4	+8.2e-4	+1.6e-2	+4.5e-2	+2.0e-2	+3.3e-4	+2.6e-2	+2.6e-3
$P_{k,2}$	+7.3e-1	+1.1e-2	+4.2e-3	+4.6e-3	+2.3e-2	+2.1e-2	+2.2e-2	+9.6e-2	+3.1e-2	+1.1e-2	+2.7e-2	+8.5e-3	+1.5e-2
$P_{k,3}$	+6.9e-1	+1.0e-2	+1.8e-2	+6.2e-3	+3.3e-2	+4.3e-3	+4.7e-3	+9.4e-2	+5.3e-3	+5.7e-2	+1.9e-2	+6.4e-3	+5.3e-2
$P_{k,4}$	+6.7e-1	+5.0e-3	+4.4e-3	+7.7e-3	+7.6e-3	+2.3e-3	+5.4e-3	+1.8e-1	+7.7e-2	+1.7e-2	+1.7e-2	+2.3e-3	+6.0e-3
$G_{k,1}$	-1.3e-1	-4.8e+0	-8.9e+0	-5.8e+0	-6.4e+0	-8.0e+0	-7.1e+0	-4.1e+0	-3.1e+0	-3.9e+0	-8.0e+0	-3.6e+0	-5.9e+0
$G_{k,2}$	-3.2e-1	-4.5e+0	-5.5e+0	-5.4e+0	-3.8e+0	-3.8e+0	-3.8e+0	-2.3e+0	-3.5e+0	-4.5e+0	-3.6e+0	-4.8e+0	-4.2e+0
$G_{k,3}$	-3.7e-1	-4.6e+0	-4.0e+0	-5.1e+0	-3.4e+0	-5.4e+0	-5.4e+0	-2.4e+0	-5.2e+0	-2.9e+0	-4.0e+0	-5.1e+0	-2.9e+0
$G_{k,4}$	-4.0e-1	-5.3e+0	-5.4e+0	-4.9e+0	-4.9e+0	-6.1e+0	-5.2e+0	-1.7e+0	-2.6e+0	-4.1e+0	-4.1e+0	-6.1e+0	-5.1e+0
a_k	-1.5e-3	-6.6e-3	+1.1e-2	+6.4e-3	-2.2e-3	-3.7e-3	+1.4e-3	+1.3e-2	+1.1e-2	-1.2e-3	+1.3e-2	-1.7e-2	-5.1e-3
b_k	-2.2e-1	-4.3e+0	-6.3e+0	-5.7e+0	-4.2e+0	-5.3e+0	-5.3e+0	-3.4e+0	-4.6e+0	-3.7e+0	-5.5e+0	-3.7e+0	-3.9e+0
R^2	+7.4e-1	+7.1e-1	+2.7e-1	+9.1e-1	+7.3e-2	+5.3e-2	+5.9e-2	+7.8e-1	+1.9e-1	+1.3e-2	+3.5e-1	+9.6e-1	+7.4e-2
T_k	$+\infty$	$+\infty$	+4.9e+2	+6.9e+2	$+\infty$	$+\infty$	+1.8e+3	+2.2e+2	+3.4e+2	+1.1e+4	+3.6e+2	$+\infty$	$+\infty$